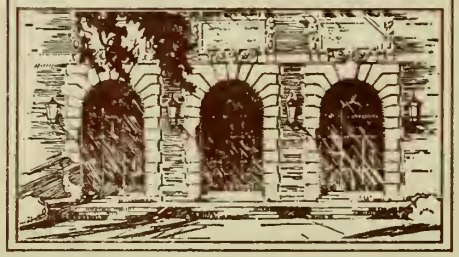
LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

331.112
H49t



The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.

To renew call Telephone Center, 333-8400

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

MAR 2 5 1981

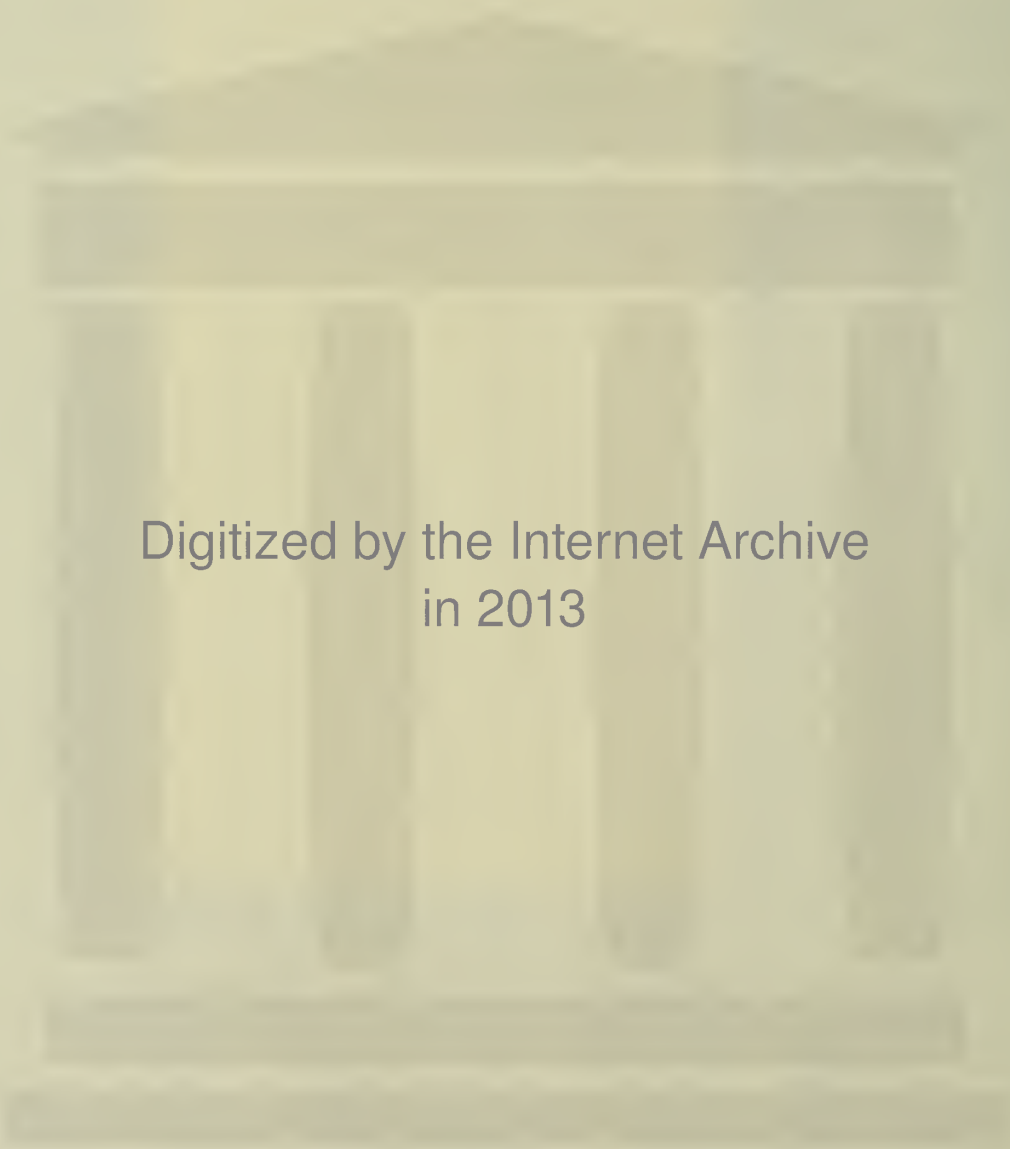
FEB 2 6 1981

DEC 1 8 1994

DEC 2 1 1994

JAN 1 3 2011

L161—O-1096



Digitized by the Internet Archive
in 2013

http://archive.org/details/towarddevelopmen00hewi

REPORT II

TOWARD THE DEVELOPMENT OF A COUNTY EMPLOYMENT PROJECTION SYSTEM FOR THE STATE OF ILLINOIS

June, 1975

A report to the

Illinois Environmental Protection Agency
Division of Air Pollution Control

by

Professor Geoffrey Hewings
Department of Geography
and
David G. Gerard
Bureau of Urban and Regional Planning Research
University of Illinois at Urbana-Champaign

ACKNOWLEDGMENTS

This report was prepared by the Bureau of Urban and Regional Planning Research, University of Illinois at Urbana-Champaign, for the Illinois Environmental Protection Agency (IEPA), Division of Air Pollution Control under Contract No. P-32, using grant funds provided by the United States Environmental Protection Agency (U.S. EPA) through its Region V office in Chicago.

The opinions, findings and conclusions expressed are those of the authors and not necessarily those of the IEPA or the U.S. EPA.

The IEPA gratefully acknowledges the assistance, suggestions and participation of staff members in the Growth and Development Division of the Illinois Bureau of the Budget in the preparation of this report.

Address Inquiries To:
Illinois Environmental Protection Agency
Division of Air Pollution Control
2200 Churchill Road
Springfield, Illinois 62706

ATTN: AQMA Coordinator
Phone: (217)-782-5811

Project Staff

Principal Investigators

Professor Andrew Isserman, Department of Urban and Regional Planning and Department of Economics

Professor Geoffrey Hewings, Department of Geography

Computer Programmers

Robert Frerichs

Charles Hohmeier

Research Assistants

Charles Cumby
Richard Edminster
David Gerard
Robert Gillespie
Margaret Martino
Luis Ortiz
Cheryl Rampke
Peter Skaperdas
Margaret Winter

Typists

Ellen Abell
Barbara Hartman
Dorinda Rice
Harold Strever

PREFACE

States are required to develop comprehensive Air Quality Maintenance Area (AQMA) plans for submission to the U.S. Environmental Protection Agency (U.S. EPA) for those areas where it is anticipated that, either because of existing air quality or projected growth, national ambient air quality standards will not be maintained through 1985 and beyond. This requirement arises out of a court decision in the U.S. Court of Appeals, District of Columbia Circuit, January 31, 1973, involving the Natural Resources Defense Council, Inc., et al, vs. EPA (475 F.2d 968 (1973)).

A key element in the development of meaningful AQMA strategies is the utilization of economic and population growth projections in the definition of future pollution problems. Historically, efforts in this area have been directed toward the correlation of demographic data with pollution levels, i.e., studying the relationship between certain data (such as population, earnings, and traffic volumes) and observed pollutant levels. The set of pollutant indicators thus established can be projected into the future using standard techniques and the probability of a pollution problem thereby established. However, scant attention has been devoted in these efforts to an intensive scrutiny of the validity and accuracy of the growth projection techniques themselves. This problem is particularly true as it relates to the disaggregation of projected population and economic growth information below the county spatial level.

In most states there are some internal constraints on the use of growth projection information for planning purposes. For example, in Illinois, the State Bureau of the Budget prepares growth projections which must be utilized, $\pm$ 5%, by all state agencies for planning purposes. However, the lowest level of spatial disaggregation of these data is, in general, currently the county for population and the multi-county region for employment. The sub-county allocation, i.e., the stepping-down of these data, is now left as an exercise for regional or county planning bodies. The methodologies utilized in stepping-down these county-level projections are not necessarily consistent from county to county, or planning area to planning area. In certain instances, well-defined procedures do not exist.

If states are to adequately identify and evaluate "hot spots" within AQMA's and promulgate substantive strategies for effectively dealing with sub-county/regional pollution problems, rational and coherent step-down procedures must be developed and utilized consistently. Such procedures will also be useful in implementing the requirements of a statewide non-deterioration policy.

The purpose of this project was to evaluate the relative accuracy of various methodologies for stepping-down county-level population and economic growth projections to a defined sub-county spatial level. Additionally, other problem areas were investigated to assist the IEPA in adequately addressing questions concerning the validity of projected future growth and development conditions as they relate to potential air pollution problems. The following reports constitute the results of the project:

Report I: Development and Testing of a Township Population Projection System for the State of Illinois by Professor Andrew Isserman

Report II: Toward the Development of a County Employment Projection System for the State of Illinois by Professor Geoffrey Hewings and David Gerard

Report III: A Survey of Projection Activities of Selected Planning Agencies in the State of Illinois by Cheryl Rampke

Report IV: Factors Affecting the Location of Selected Industries Subject to Federal New Source Performance Standards: A Primer by students in the Economic Analysis Workshop, edited by Professor Andrew Isserman

Illinois Environmental Protection Agency
Division of Air Pollution Control
October, 1975

TABLE OF CONTENTS

		Page
Section I:	INTRODUCTION	1
Section II:	THE EMPLOYMENT SECTORS AND MODELS	4
Section III:	PREVIOUS RESEARCH IN THE AREA OF COUNTY EMPLOYMENT PROJECTIONS	12
Section IV:	THE ILLINOIS RESULTS	19
Section V:	SUMMARY AND CONCLUSIONS	33
Section VI:	NEW MODEL DESIGN	35
Appendix:	MORE SOPHISTICATED MODELS FOR SINGLE AREA ANALYSIS	52

I. INTRODUCTION

The vast majority of the employment step-down models resembles the set of models used for the population component of this study. While these models, and simple variations of them, have been utilized in a large number of studies, the research reported here represents the first time that a sizable family of models has been applied to one region. Hence, the analysis takes on a further dimension of interest in that results obtained from each model are directly comparable since the same data base was used. The results, however, are by no means encouraging in the context of scientific acceptability: none of the models provided estimates that were within acceptable limits or consistent over sets of counties or industrial sectors. This is one particular area in which additional research would appear to be mandatory.

This Report is divided into the following sections. The models, employment categories and the tests adopted are discussed in the next section. There follows a brief guide to previous research in the area of county employment projections, with particular emphasis on the accuracy that other authors have been able to achieve. The Illinois results are evaluated next, followed by some conclusions. The final section presents some thoughts on new model design, with the caveat that, wherever possible, the models should require as little additional data as possible. The Appendix to this Report documents a somewhat different approach to substate population and employment analysis. A guide and critical evaluation are given of single region models, from economic base analysis to land use allocation models, with particular emphasis on the interaction of economic and ecologic systems in the situations in which such analysis has been undertaken or suggested. Thus, the Appendix would provide guidance on the selection of an appropriate model for the analysis of economic activities within a single region--for example, a Standard Metropolitan Statistical Area.

TABLE 1

EMPLOYMENT SECTORS USED IN THE PROJECTION MODELS

Project Number	SIC Code	Brief Description	Source Data
1	10	Metal Mining	Illinois Bureau of Employment Security Records 1st Quarter Each Year
2	11	Anthracite Mining	
3	12	Bituminous Coal and Liguite Mining	
4	13	Crude Petroleum and Natural Gas	
5	14	Mining, Quarrying Non-Metallic Minerals	
6	19	Ordinance and Accessories	
7	20	Food and Kindred Products	
8	21	Tobacco Manufactures	
9	22	Textile Mill Products	
10	23	Apparel	
11	24	Lumber and Wood Products (ex. Furniture)	
12	25	Furniture and Fixtures	
13	26	Paper and Allied Products	
14	27	Printing, Publishing and Allied Industries	
15	28	Chemical and Allied Products	
16	29	Petroleum Refining	
17	30	Rubber and Miscellaneous Plastic Products	
18	31	Leather and Leather Products	
19	32	Stone, Clay, Glass and Concrete Products	
20	33	Primary Metal Industries	
21	34	Fabricated Metal Products	
22	35	Machinery (ex. Electrical)	
23	36	Electrical Machinery	
24	37	Transportation Equipment	
25	38	Professional Products	
26	39	Miscellaneous Manufacturing	
27	--	Construction	
28	--	Transportation, Construction, Public Utilities	BEA
29	--	Trade	
30	--	Fire, Insurance, Real Estate	
31	--	Services	
32	--	Others, n.e.s.	

TABLE 2

MULTI-COUNTY REGIONS: ILLINOIS

Region 1	Region 2	Region 3	Region 4
McHenry	Iroquois	Douglass	Jasper
Lake	Ford	Edgar	Crawford
Kane	Vermillion	Coles	Richland
Cook	Champaign	Clark	Lawrence
DuPage	Piatt	Cumberland	Wayne
Kendall			Edwards
Will			Wabash
Grundy			White
Kankakee			

Region 5	Region 6	Region 7	Region 8
JoDaviess	Bureau	Livingston	Macon
Stephenson	LaSalle	McLean	Moultrie
Winnebago	Putnam	DeWitt	Shelby
Boone			
Carroll			
Ogle			
DeKalb			
Lee			

Region 9	Region 10	Region 11	Region 12
Fayette	Jefferson	Stark	Cass
Effingham	Perry	Marshall	Menard
Marion	Franklin	Peoria	Logan
Clay	Hamilton	Woodford	Scott
	Jackson	Fulton	Morgan
	Williamson	Tazewell	Sangamon
	Saline	Mason	Christian
	Gallatin		Macoupin
	Union		Montgomery
	Johnson		
	Pope		
	Hardin		
	Alexander		
	Pulaski		
	Massac		

Region 13	Region 14	Region 15
Madison	Whiteside	Hancock
Bond	Rock Island	McDonough
St. Clair	Henry	Adams
Clinton	Mercer	Schyler
Monroe	Henderson	Brown
Randolph	Warren	Pike
Washington	Knox	
Greene		
Calhoun		
Jersey		

II. THE EMPLOYMENT SECTORS AND MODELS

Employment Sectors

The list of employment sectors used in this analysis is shown in Table 1. For the most part, the sectors represent a division of employment into what are referred to as two-digit Standard Industrial Classification (SIC) groups. These are broad groups of activities: for example, SIC 20 Food and Kindred Products would include SIC 201 Meat Products; SIC 202 Dairy Products; SIC 203 Canned, Cured and Frozen Goods; SIC 204 Grain Mill Products; and so forth. These three-digit subgroups of SIC 20 are themselves further subdivided into four-digit groups (for example, SIC 2011 Meatpacking Plants). The sectors which are not numbered are aggregates of two or more two-digit SIC groups. Construction would include SIC 15, 16 and 17; Trade would include SIC 50 through 59; Finance, Insurance and Real Estate comprises SIC groups 60 through 67.

The employment data were available for the following years: 1959, 1960, 1962, and 1965 through 1972. A larger set of data was available for the years 1967 through 1972: these data included the farming sectors, federal and state employment and federal military employment. Inasmuch as the objective of the analysis was to test five-year projections, it was not possible to utilize these data. Thus, the models were tested for employment sectors which account for approximately three-quarters of total state employment. Of greater importance is the fact that the sectors tested include the majority of those activities that are likely to be the major point source polluters. All the models, which are described below, are used as though 1972 was the projection year: the set of data through 1967 was used to "forecast" employment levels in 1972. On this basis, tests of the accuracy of these forecasts could be made.

The Models

To assist the description of the models, the following notation will be used. Let:

i represent the county (i=1.....102) and $i \varepsilon j$

j represent the multi-county (j=1....15)

k represent the employment sector (k=1....32)

t represent the time period

n represent the number of years in the future (projection period, which in this case is 5 years)

* represent a projected value

** represent a projected value adjusted such that the sum of the county estimates conforms to the multi-county total in each employment sector

Model 1: Average Rates, Exponential Growth

For each employment class k in county i ($i \varepsilon j$), the annual rate of growth is calculated as follows:

$$r^{ijk}_{t+1} = E^{ijk}_{t+1} \; / \; E^{ijk}_{t} \qquad \text{for } t=1959,\ 1965,\ 1966$$

$$r^{ijk}_{1962} = \sqrt{E^{ijk}_{1962} \; / \; E^{ijk}_{1960}}$$

$$r^{ijk}_{1965} = \sqrt[3]{E^{ijk}_{1965} \; / \; E^{ijk}_{1962}}$$

The reason for different computation relates to the fact that annual data are not available for each year between 1959 and 1967: consequently, adjustments have to be made to transform two- and three-year growths to annual growth rates.

The average annual growth rate is then computed:

$$\bar{r} = \frac{\sum_{t=1}^{m} r^{ijk}_{t+1}}{m} \qquad \text{where m is the number of years of observation}$$

The projected employment level n years ahead is obtained from:

$$E^{ijk*}_{t+n} = (\bar{r}^{ijk})^{n} \cdot E^{ijk}_{t}$$

Adjustment to multi-county j's independently projected employment total in sector k is made by summing the separate county projections to provide an estimate of the multi-county total:

$$E^{jk*}_{t+n} = \sum_{i} E^{ijk*}_{t+n} \qquad (i \varepsilon j)$$

Then, the adjusted county employment is derived from:

$$E^{ijk**}_{t+n} = [E^{jk}_{t+n} / E^{jk*}_{t+n}] . \; E^{ijk*}_{t+n}$$

where E^{jk}_{t+n} is the independently forecast multi-county employment in sector k.

Model 2: Average Rates, Linear Growth

For each employment class k in county i, growth is calculated:

$$r^{ijk}_{t+1} = E^{ijk}_{t+1} - E^{ijk}_{t}$$

The mean is computed by

$$\bar{r}^{ijk} = \frac{\sum_{t=1}^{m} r^{ijk}_{t+1}}{m}$$

where m is the length of the employment record

The employment n years ahead is obtained from:

$$E^{ijk*}_{t+n} = E^{ijk}_{t} + [\bar{r}^{ijk}] . \; n$$

This first estimate is then adjusted to the multi-county totals in the same manner as that described for Model 1.

Model 3: Regression Rates, Exponential

For each employment class k in county i, the following equation is estimated using ordinary least squares (OLS) regression:

$$\log E^{ijk}_{t} = a + b^{ijk}(t)$$

The antilog of the coefficient of t is r^{ijk}: the first estimate of county employment in k is obtained by

$$E_{t+n}^{ijk*} = [r^{ijk}]^n \cdot E_t^{ijk}$$

Similar procedures to the one outlined in Model 1 were used to adjust these estimates to conform with the multi-county totals.

Model 4: Regression Rates, Linear

In this model, the OLS form was as follows:

$$E_t^{ijk} = a + b^{ijk}(t)$$

Then,

$$E_{t+n}^{ijk*} = E_t^{ijk} + b^{ijk} \cdot n$$

The adjustment process is repeated here. Note that the regression models forecast using the slope of the line from the last point of observation rather than extrapolating the whole regression line n years into the future.

Model 5: Constant Share

Define a county share of employment sector k as follows:

$$r^{ijk} = E_t^{ijk} / E_t^{jk}$$

Then,

$$E_{t+n}^{ijk*} = r^{ijk} \cdot E_{t+n}^{jk}$$

Model 6: Projected Share, Exponential

For each employment sector k in county i, define r_t^{ijk} as in Model 5 but for each time period t. The logarithm of this share is then regressed against t. Let the antilog of the coefficient of t be c^{ijk}.

The forecast share is then defined as:

$$r_{t+n}^{ijk*} = [c^{ijk}]^n \cdot r_t^{ijk}$$

Define:

$$R^{jk}_{t+n} = \sum_i r^{ijk*}_{t+n}$$

Then:

$$E^{ijk*}_{t+n} = [r^{ijk*}_{t+n} \; / \; R^{jk}_{t+n}] . \; E^{jk}_{t+n}$$

In this case, the forecast shares are implicitly adjusted so that the total employment in the multi-county region in sector k is the sum of the projected shares in each of the constituent counties. This is effected by making the forecast share for the individual county a percentage of the total forecast shares of the counties within the multi-county region.

Model 7: Projected Share, Linear

In this model, r^{ijk}_t is defined as for Model 6; these values are then regressed against t (rather than log r^{ijk}_t against t). Let b^{ijk} be the regression coefficient of t, then:

$$r^{ijk*}_{t+n} = r^{ijk}_t + b^{ijk} . \; n$$

R^{jk}_{t+n} is defined in the same manner as in Model 6 and the same procedures are followed to arrive at a value for E^{ijk*}_{t+n}.

Model 8: Simple Shift-Share

Define the following rates of growth for each employment sector k:

$$r^{jk} = E^{jk}_t \; / \; E^{jk}_{t-5}$$

$$r^{ijk} = E^{ijk}_t \; / \; E^{ijk}_{t-5}$$

$$r^{jk*} = E^{jk}_{t+5} \; / \; E^{jk}_t$$

Then:

$$E_{t+5}^{ijk*} = [r^{jk*} + (r^{ijk} - r^{jk})] . E_t^{ijk}$$

In this model, it is assumed that the differential growth rate between the county and the multi-county for sector k between time period t-5 and t will be maintained between t and t+5. Thus, county employment growth in sector k is considered to be composed of two elements: (1) the share $[r^{jk*}]$ and (2) the "shift" $[r^{ijk} - r^{jk}]$. The latter term is really a misnomer since employment does not necessarily move from one part of space to another. What it implies is best described by its synonomous term, competitive component. Thus, if in the previous time period, employment in k grew at the same rate in the county as in the multi-county, the competitive component would be zero--indicating that the sector in the county was at neither an advantage or a disadvantage vis a vis the multi-county. A negative competitive component would indicate less rapid growth in the county than the multi-county, while a positive value would imply the reverse. Thus the competitive component may take on a positive, negative or zero value.

Model 9: OBERS Shift-Share

This model could also be regarded as a double exponential version of the projected share exponential model (Model 6). It is referred to as a shift-share model because the competitive component is calculated implicitly rather than explicitly as in the case of Model 8. For each employment sector k in county i, the following regression is obtained:

$$\log [E_t^{ijk} / E_t^{jk}] = \log a + b^{ijk} \log t$$

or, in the notation of Model 6:

$$\log r_t^{ijk} = \log a + b^{ijk} \log t$$

Then:

$$E^{ijk*}_{t+n} = \text{antilog}\,[\log (E^{ijk}_{t} \;/\; E^{jk}_{t}) + b/t+n]\,.\; E^{jk}_{t+n}$$

The term in the square brackets provides the estimate of the logarithm of the share of multi-county employment in sector k that will be found in county i. Multiplying the antilog of this by the independently projected value of multi-county employment will yield the county employment value at t+n.

Model 10: Population/Employment Projection

Let P^{ij}_{t} and P^{j}_{t} refer to population totals in the county and multi-county respectively in time period t. Define:

$$r^{jk} = E^{jk}_{t} \;/\; E^{jk}_{t-10}$$

$$r^{ij} = P^{ij}_{t} \;/\; P^{ij}_{t-10}$$

$$r^{j} = P^{j}_{t} \;/\; P^{j}_{t-10}$$

Then:

$$E^{ijk*}_{t} = r^{jk}.\; E^{ijk}_{t-10}\;.[r^{ij} \;/\; r^{j}]$$

In this model, the estimated county employment in sector k is assumed to remain the same share of multi-county employment with a modification based on the relative rates of growth in multi-county and county populations. Procedures used in the earlier models were adopted to ensure that the sum of county estimates was in accord with the multi-county total employment in each sector. In the application in this model, the employment estimates were made for 1972, modified by the relative population changes between 1960 and 1970.

Model 11: Simple Population/Employment Extrapolation

This very simple model projects county employment in sector k as follows:

$$E^{ijk*}_{t+n} = [P^{ij}_{t+m} / P^{j}_{t+m}] . E^{jk}_{t+n}$$

where the population values are for the latest available year and $m<n$. In the application to the Illinois data, county employment estimates were adjusted by the relative population ratio for 1970.

Model 12: Modified Shift-Share

Define the following growth rates:

$$r^{jk} = E^{jk}_{t+5} / E^{jk}_{t-5}$$

$$r^{jk*} = E^{jk}_{t+5} / E^{jk}_{t}$$

Then:

$$E^{ijk*}_{t+5} = r^{jk*} . E^{ijk}_{t} - 0.5\ r^{jk} . E^{ijk}_{t-5}$$

In this model, the changing share of county employment between t-5 and t is extrapolated to t+5. The second term on the right-hand side is multiplied by 0.5 since r^{jk} represents a growth over ten years (t-5 to t+5) whereas r^{jk*} is calculated over only a five-year period (t to t+5).

III. PREVIOUS RESEARCH IN THE AREA OF COUNTY EMPLOYMENT PROJECTIONS

Previous Work

It was noted in the Introduction to this Report that this present study represents the first time that a large number of the more popular models used for making projections have been applied to a common data base. Perhaps, the most recent study that compared the results of a number of different models was that conducted by Greenberg (1972) using data for the New York Metropolitan Region for the years 1947, 1958 and 1963 (the forecast year). In addition, population data were used in some of Greenberg's models, the application being made in the context of models like those described as Model 10 and 11 in this Report. However, with a three-point observation data base, Greenberg was not able to test any linear or exponential share or regression models.

The objective of Greenberg's study was to estimate value added for forty-two industries for twenty-three counties within the New York Region for five-year periods between 1970 and 1985. The analysis was conceived as a subsystem input into a broader study of water supply and waste disposal in the region. The procedure involved allocating national projections to the New York Metropolitan Region and then allocating these projections to the constituent counties of the Region. The procedure is similar to the one in hand for Illinois although the Illinois Bureau of the Budget is responsible for the estimation of the multi-county qua regional estimates.

At the regional level (i.e., stepping down national projections to the New York region), Greenberg (1972) was able to obtain mean errors of between 11 and 26 percent. In Table 3, the mean percentage errors are shown for a selection of SIC groups "forecast" by Greenberg and compared to actual 1967 data at the county level. Greenberg's model 1 involved a regression analysis of the more classical variety (as opposed to its application to the Illinois data in this Report). Relative changes in county output between 1958 and 1963 were related

TABLE 3

COUNTY PROJECTIONS OF EMPLOYMENT FOR NEW YORK, 1967
MEAN PERCENTAGE ERRORS

SIC Group	Model 1 (Regression)	Model 2 (Constant Share)	Model 3 (Implicit Shift Share)	Model 4 (Linear Extrapolation 1958-1963 Output)
20	18.4	22.2	19.4	20.4
22	--	14.7	15.4	18.4
26	10.8	12.5	15.6	15.2
28	--	14.2	16.6	15.9
31	50.2	27.4	66.2	63.8
33	22.5	25.3	49.5	47.7

SOURCE: Greenberg (1972, p. 413).

to the following set of variables: (1) gross population density, 1963; (2) change in employment in manufacturing 1958-1963; (3) land zoned for industry, 1962; and (4) port, rail, water and sewer service, 1962. Suitable equations were obtained only for those SIC groups shown in Table 3. Greenberg's model 2 and model 3 are analogous to the present study's Models 5 and 12, while the fourth Greenberg model may be considered the same type as Model 7 in this study.

Models 3 and 4 in this study have been utilized in the SELUM model developed by Argonne National Laboratory (1973). However, published results for only a small number of counties were included in the Appendix to their report, precluding any direct comparison with the Argonne results. This exercise would have been useful since the Argonne manufacturing data were assembled from the Illinois Manufacturers' Directory (IMD) rather than the Census of Manufactures. A brief comparison with Argonne estimates for Winnebago County for the years 1965, 1967, 1969 and 1970 revealed rather significant discrepancies in employment estimates for the two-digit SIC sectors present in that county. In fairness, it should be pointed out that the IMD data were acknowledged by Argonne to be incomplete

> Because the inventory (IMD) was constructed from a voluntary subscription (and thus) its contents should be checked for consistency and accuracy with other data (if available) prior to its use as a basis for obtaining projections. [Argonne (1973) p. 51]

The OBERS Shift Share model (Model 9) was developed for projections of regional economic activity in the US by a joint research group from the Office of Business Economics (now Bureau of Economic Analysis) and the Economic Research Service of the US Department of Commerce. The OBERS (1972) models have not been subjected to empirical testing at the county level as of the present time.

Several other models have been developed and tested representing variations on shift and share analysis. These would include the work of Brown (1969), James and Hughes (1973), Stillwell (1969), Ashby (1970), Chalmers (1971) and Houston (1967). Most of the analysis, however, has been at the regional, state or SMSA level. Even here, there is considerable debate about the utility of these models for predicting future activity levels. A number of other authors have attempted to introduce elements of location theory, notions of agglomeration or lag effects into their models (see, for example, Hellman and Marcus (1970a, 1970b), Hellman (1974), Czamanski (1964, 1965) and Chalmers and Beckhelm (1974)). Most of these models require considerable data inputs beyond employment by sector: some of them are discussed in greater detail in the section on new model design (Section VI).

General Difficulties Associated with Employment Projection at the County Level

It is quite obvious that as one moves from the national to the regional to the county level, the mean percentage errors increase. This is not really unexpected for a number of reasons. First, the magnitude of employment in any sector is likely to be smaller and hence, minor changes in employment levels will be represented as correspondingly larger percentages changes at the county level than at the national level. Secondly, national business cycle movements may not be felt with the same degree of variability at the county level as at the national or regional/multi-regional level. Thus, it may well be possible for some counties to experience very little disruptive effects in some sectors during downturns in business activity whereas the aggregate effects at the multi-county level may be rather severe. In addition, the time incidence of business cycle impacts in the county may not parallel those at the multi-county or national level. Minor variations in lead and lag effects between

the county and the multi-county would thus undermine the accuracy of models based on assumptions about constancy of county shares of multi-county employment in any given sector.

Thirdly, a related problem to the one just noted arises, this time in the context of the years for which data are available. The Illinois data used in this present analysis encompass the years from 1959 through 1972 (with the exceptions noted earlier). The beginning of the data record marks the entry of the national economy into a major recession: during the middle period, the economy, was, for the most part, enjoying unparalled prosperity while the end of the data record once again marks the return of the economy into recession. Thus, it becomes very difficult to determine whether the inability of a model to accurately forecast employment levels by sector arises from inherent conceptual difficulties with the model itself or a set of data that reflects a period of marked fluctuation in business activity. Unfortunately, the vicissitudes of economic health suggest that the search for a "pure" period of economic data would be fruitless. The models have to operate in the context of a system in which there would appear to be a large measure of uncertainty.

A fourth reason for difficulty in making county allocations relates to the issue of product and industry mix. When the SIC code was described briefly at the beginning of Section II, it was noted that each industrial sector is comprised of a set of firms that are producing products which have some affinity but do not necessarily involve the same types or quantities of inputs, or the same quality and quantity of labor and capital. Furthermore, these firms may not sell to the same customers. Unless the mix of firms within any of the two-digit sector at the county level mirrors that at the multi-county level, it is quite probable that fluctuations in employment will be different, responding as they do to different sets of forces. Thus, increases in employment in one sector in

one county at a time when employment is steady or declining in the same sector in other counties within the multi-county region may be indicative of industry mix differences rather than any comparative advantage. While three- and four-digit SIC data are available at the county level from the County Business Patterns, the record is not complete since disclosure problems preclude the revelation of certain industry data when the number of firms or the total employment fall below certain federal norms.

A final problem involves the prediction of firm closure or the opening of new establishments: in the aggregate, that is at the national or state level, models may be developed to predict the probability of new industries locating in a state or the probability of firm closure. However, the problem becomes virtually intractable when specific detail (the industrial sector) is required within the context of county spatial units. It was found that, in some cases, the sudden closure or opening of firms contributed as much as 7 percent to the mean percentage errors in the county aggregate employment forecasts (see next section). The reduction in errors, however, could not be completed since identification of new activity could only be made when zeros in one year were replaced by employment levels greater than zero in following years or when the reverse occurred. When values greater than zero appeared in all years, it was obviously not possible to trace how much the fluctuation in employment levels was due to incumbent firm expansion or contraction and how much was due to new firm location or firm withdrawal. Again, some measure of this latter effect could be gained from an analysis of the County Business Patterns, wherein the number of establishments is listed for each sector.

Quite clearly, there are many reasons why one should not be surprised to find larger mean percentage errors in the employment forecasts than in the population forecasts. These factors notwithstanding, some of the results show

promise for future endeavors even if the overall picture is somewhat disappointing.

The models chosen for testing at this time represent a set of the more popular ones currently used by planning agencies at various levels. As a result, they represent a set which is currently possible to implement empirically rather than a set which also includes some of the translations of recent theoretical developments in regional and industrial growth and development theories into empirical implementation. Because they are simple to operate, have modestly avaricious data requirements and reflect, in part, some of the tried (if not necessarily true) beliefs about growth relationships from one level in the system to the next lowest, they are the ones which were calibrated for the Illinois data. In the next section, many, many pages of computer output and data analysis have been condensed into a sample of results which highlight the main findings rather than a complete analysis of every operation that was performed.

IV. THE ILLINOIS RESULTS

A Brief Note on the Data Used and the Testing Procedure

If one refers back to Table 1, one will notice that 26 of the 32 employment sectors were obtained from the Illinois Bureau of Employment Security Records (ES202): the remaining sectors were obtained from the Bureau of Economic Analysis county records. The former set of data relate to all covered employment in the state: the data record for each county records employment by SIC sector for the first quarter of each year. The BEA data are obtained, in part, from ES202 records and in part from other state data, national industry estimates and the like.

Inasmuch as the set of data for Illinois represents a complete "population" rather than a sample, standard statistical techniques were not utilized in the analysis of the errors. It was felt that knowledge of mean percentage errors and its distribution spatially (i.e., by county) and sectorally (i.e., by employment sector) would be more useful than measures such as Theil's test, one of the tests which has been most prominently visible in the literature.

The Results

Summaries of selected aspects of the data analysis are given in Tables 4 through 11. From these tables, it becomes evident very quickly that the magnitude of the errors, in comparison to the population forecasts, are of the order of five or six times greater in terms of mean percentage errors. However, for some sectors and for some counties, the estimates were almost as good as those obtained in the population analysis: sadly, the number of sectors and counties in this category was very small!

A. State Estimates of Total Employment

The state totals, representing the mean of the mean percentage errors for each of the 15 multi-county regions, are shown in Table 6. Note that the best

model, Model 5 (constant share), achieved a mean percentage error of in excess of 54 percent, while the worst model, Model 11 (the simple population/employment projection) provided an error several times greater in magnitude--304 percent! The variability between the models was much narrower, however, with nine of the models achieving mean percentage errors of between 54 and 84 percent. The remaining three had errors in excess of 100 percent. These errors are somewhat misleading in that each employment sector in each county is weighted equally, irrespective of the magnitude of the employment total in that sector. Hence, as happened with some models, overestimates of 3000 percent were obtained in some sectors at the county level in sectors with total employment of only 100 in the county. On the other hand, an overestimate of 15 percent in another county may represent 2000 employees in that particular sector. Where appropriate, especially at the county level, the average number of employees over or under estimated is also shown together with the mean percentage error. This will allow the reader to judge more carefully the efficacy of a particular model.

Whatever the frailties of the mean percentage error as a summary statistic, it does point out that, in the aggregate, certain models cannot be considered as suitable candidates for stepping down multi-county projections. Models 1, 8 and 11 would seem to fall into that category. However, the state totals are an inaccurate guide to model performance on a sector by sector and county by county basis.

B. State Estimates by Employment Sectors

The state summaries for individual sectors are shown in Table 4: the best estimates and the model providing them are shown in Table 5 together with an indication of the errors expressed in terms of number of employees over or underestimated rather than mean percentage errors. From these tables, the discarding of Models 1, 8 and 11 would still seem to be justified since none of these models

TABLE 4

STATE SUMMARIES FOR INDIVIDUAL SECTORS, 5 YEAR PROJECTION, 1967-1972
MEAN PERCENTAGE ERRORS

Employment Sector	Model 1	2	3	4	5	6	7	8	9	10	11	12
10	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.043	0.029	0.029	0.347	0.029
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.402	0.423	0.390	0.404	0.345	0.371	0.384	0.656	0.343	0.671	0.683	0.430
13	0.982	1.080	1.185	1.229	0.624	1.008	1.065	2.085	1.016	0.757	2.580	1.025
14	1.344	0.531	0.520	0.498	0.415	0.514	0.488	0.794	0.442	0.746	0.900	0.515
19	0.222	0.222	0.204	0.116	0.222	0.219	0.219	2.000	0.222	0.286	0.195	0.286
20	5.094	0.657	0.665	0.643	0.544	0.671	0.661	0.797	0.571	1.246	5.299	0.649
21	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.003	0.223	0.003
22	0.317	1.493	0.337	0.655	0.269	0.337	0.338	1.212	0.317	0.322	0.321	0.387
23	0.789	0.645	0.620	0.641	0.561	0.658	0.696	0.848	0.707	0.834	6.141	0.614
24	1.311	0.882	0.750	0.717	0.567	0.737	0.708	1.433	0.867	1.164	3.515	0.876
25	2.131	0.787	0.996	0.762	0.610	1.042	0.862	2.915	0.846	0.596	4.304	0.931
26	0.832	0.304	0.480	0.273	0.293	0.495	0.325	0.913	1.273	1.450	0.660	0.370
27	0.733	0.359	0.419	0.347	0.334	0.421	0.385	0.536	0.329	0.521	2.180	0.438
28	0.884	0.518	0.732	0.500	1.182	0.788	0.864	2.161	1.206	1.991	1.574	1.023
29	0.306	0.343	0.274	0.286	0.297	0.279	0.286	0.958	0.406	0.425	2.058	0.335
30	2.652	2.005	2.105	2.005	2.001	2.149	2.048	2.203	0.851	1.176	2.887	2.042
31	0.371	0.293	0.235	0.258	0.205	0.237	0.220	0.421	0.204	0.488	2.631	0.261
32	1.505	0.949	0.770	0.833	0.725	0.809	0.780	1.153	0.768	1.050	4.744	0.837
33	1.783	0.783	0.724	0.670	0.554	0.745	0.723	1.170	0.599	0.753	3.757	0.660
34	0.958	1.004	0.789	0.790	0.634	0.809	0.825	1.089	0.609	1.032	7.402	0.863
35	1.394	0.794	0.908	0.811	0.822	0.941	0.869	1.976	0.972	1.263	9.399	0.925
36	1.673	0.722	1.077	0.659	0.644	1.056	0.671	1.700	0.698	1.074	3.784	0.791
37	0.665	0.555	0.491	0.520	0.594	0.542	0.577	1.384	0.736	1.439	3.656	0.792
38	4.575	0.901	1.115	0.898	0.742	1.116	0.924	1.290	0.349	0.685	3.267	0.946
39	1.472	0.425	0.486	0.426	0.436	0.509	0.470	0.917	0.572	1.082	1.998	0.522
Construction	2.121	0.497	0.736	0.430	0.408	0.735	0.522	2.034	0.447	0.406	0.963	0.716
Transportation	1.614	0.195	0.215	0.167	0.144	0.226	0.214	0.677	0.240	0.294	0.831	0.287
Trade	0.572	0.134	0.164	0.119	0.112	0.164	0.140	0.302	0.113	0.167	0.364	0.181
Finance	0.361	0.220	0.227	0.192	0.176	0.226	0.219	0.329	0.178	0.221	0.860	0.269
Services	1.899	0.128	0.307	0.128	0.129	0.305	0.197	1.749	0.192	0.410	0.603	0.313
Other	1.605	1.121	1.108	1.094	1.097	1.186	1.208	3.169	1.041	1.362	1.273	1.246

TABLE 5

BEST STATE ESTIMATES FOR EMPLOYMENT SECTOR, 5 YEAR PROJECTION, 1967-1972

Employment Sector	Model Number*	Average Difference**	Mean Percentage Error
10	1, 2, 3, 4, 5, 6, 7, 9, 10, 12	1	0.029
11	--***	--	--
12	9	159	0.343
13	5	12	0.624
14	5	20	0.415
19	4	2	0.116
20	5	99	0.544
21	1, 2, 3, 4, 5, 6, 7, 9	1	0.002
22	1, 9	40	0.317
23	5	41	0.561
24	5	21	0.567
25	10	65	0.596
26	4	82	0.273
27	9	66	0.329
28	4	233	0.500
29	3	42	0.274
30	9	232	0.851
31	9	44	0.204
32	5	46	0.725
33	5	115	0.554
34	9	144	0.609
35	2	238	0.794
36	5	350	0.644
37	3	247	0.491
38	9	116	0.349
39	2	59	0.425
Construction	10	200	0.406
Transportation	5	152	0.144
Trade	5	650	0.112
Finance	5	132	0.176
Services	2	280	0.128
Other	9	37	1.041

* See text.
** Number of employees rounded to nearest whole number.
*** No employment in the state.

TABLE 6

MULTI-COUNTY SUMMARIES, 5 YEAR PROJECTION, 1967-1972

Multi-County	Model 1	2	3	4	5	6
1	2.674	0.534	0.698	0.424	0.366	0.695
2	6.091	0.571	0.662	0.567	0.418	0.676
3	1.735	0.631	0.723	0.610	0.597	0.726
4	0.989	0.632	0.657	0.618	0.640	0.691
5	1.095	0.615	0.755	0.589	0.545	0.766
6	1.918	0.844	1.193	0.727	0.623	1.208
7	0.631	0.348	0.288	0.316	0.288	0.353
8	1.312	0.894	0.862	0.773	0.668	0.875
9	0.633	0.907	0.707	0.777	0.680	0.673
10	1.323	0.812	0.790	0.762	0.875	0.811
11	1.314	0.495	0.661	0.541	0.467	0.724
12	1.435	0.677	0.559	0.588	0.464	0.555
13	1.273	0.500	0.522	0.491	0.363	0.511
14	1.033	0.708	0.638	0.645	0.575	0.668
15	0.648	0.388	0.379	0.362	0.347	0.388
State	1.567	0.634	0.665	0.588	0.545	0.679

Multi-County	Model 7	8	9	10	11	12
1	0.413	1.240	0.410	0.527	2.039	0.477
2	0.641	1.135	0.560	0.633	1.591	0.651
3	0.693	1.739	0.641	0.921	2.740	0.785
4	0.657	1.559	0.664	1.538	4.566	0.776
5	0.646	1.216	0.541	0.706	4.697	0.702
6	0.792	3.705	0.799	0.641	6.884	0.982
7	0.368	0.882	0.470	0.570	1.715	0.425
8	0.816	1.737	1.078	0.761	4.170	1.020
9	0.690	1.020	0.581	0.725	1.631	0.887
10	0.850	1.689	0.890	1.298	2.066	1.002
11	0.669	1.248	0.497	0.614	2.904	0.538
12	0.586	1.119	0.454	0.643	2.126	0.628
13	0.481	1.209	0.524	0.881	2.418	0.485
14	0.697	1.260	0.620	0.808	6.535	0.755
15	0.395	0.845	0.411	0.536	1.856	0.413
State	0.630	1.381	0.605	0.848	3.044	0.698

provides the best available estimate. As one would have expected from the discussion of previous research in this area, the range of errors is much greater when one deals with the disaggregated employment estimates. Table 4 should be considered in conjunction with Table 8 which shows the distribution of employment, in percentage terms, by sector for Illinois in 1967. Hence, large mean percentage errors in sectors 10 through 14 are not as serious, when translated into numbers of employees, as similar errors in sectors 20, 35, 36 Trade or Services. Model 5 (constant share) and Model 9 (OBERS shift share) outperform the rest of the models by a wide margin. Some of the models achieve reasonable accuracy (less than 15 percent mean percentage errors) in the nonmanufacturing sectors.

C. Multi-County Summaries

Table 6 shows the ability of the models at the multi-county level--in other words, the summation of estimates for the individual counties grouped into the 15 regions. The best and worst results are shown in Table 7. Once again, Model 11 appears as a definite noncandidate for acceptance. In fairness, it should be noted that the model was not designed for all sectors in an economy. Greenberg (1972, p. 408) proposed the model for those sectors whose growth would be a function of the growth of the local market, the proxy for which was suggested to be population growth. On the positive side, Models 5 and 9 dominate the selections of models which perform best at this level. There would appear to be no systemmatic spatial variation in the estimates: Model 5 provided the best aggregate estimates for multi-county 1 (the most urban area in the state), multi-county 15 (a mixture of prosperous and less prosperous agriculturally-oriented counties) and multi-county 2 (an agricultural county with large urban centers).

TABLE 7

BEST AND WORST ESTIMATES OF TOTAL EMPLOYMENT:
MULTI-COUNTY SUMMARIES

Multi-County	Best		Worst	
	Mean Percentage Error	Model Number	Mean Percentage Error	Model Number
1	0.366	5	2.674	1
2	0.418	5	6.091	1
3	0.597	5	2.740	11
4	0.618	4	4.566	11
5	0.541	9	4.697	11
6	0.623	5	6.884	11
7	0.288	3, 5	1.715	11
8	0.668	5	4.170	11
9	0.581	9	1.631	11
10	0.762	4	2.066	11
11	0.467	5	2.904	11
12	0.454	9	2.126	11
13	0.363	5	2.418	11
14	0.575	5	6.535	11
15	0.347	5	1.856	11
State	0.545	5	3.044	11

TABLE 8

PERCENTAGE DISTRICUTION OF EMPLOYMENT BY SECTOR, 1967

Employment Sector	Percentage Distribution
10, 11, 12, 13, 14	0.5
19	(less than 0.1%)
20	2.4
21	(less than 0.1%)
22	0.1
23	0.7
24	0.2
25	0.5
26	0.8
27	2.2
28	1.1
29	0.2
30	0.7
31	0.2
32	0.7
33	2.2
34	2.9
35	4.6
36	4.3
37	0.9
38	0.9
39	0.7
Construction	3.9
Transportation, Communications, Public Utilities	5.8
Trade	18.5
Finance, Insurance, Real Estate	4.5
Services	14.1
Other	0.1
Total In State	4,827,278
Total Covered by Models	3,698,972 (76.6%)

TABLE 9

AGGREGATE RESULTS FOR SELECTED ILLINOIS COUNTIES, 5 YEAR PROJECTION, 1967-1972
[Model Number Shown in Parentheses]

County Group	Best Model		Worst Model	
	Mean Percentage Error	No. Employees	Mean Percentage Error	No. Employees
Most Urban				
Cook	0.083 (4)	1707	0.294 (1)	9978
Most Suburban				
DuPage	0.386 (5)	1368	2.357 (1)	5423
Second Most Urban/Suburban				
Will	0.280 (9)	393	3.723 (1)	2166
Madison	0.184 (5)	150	0.535 (11)	700
Agriculture with Large Urban Covers				
Champaign	0.264 (5)	202	2.565 (11)	403
Peoria	0.172 (5)	133	0.422 (8)	688
Wealthier Predominantly Agric.				
Kendall	0.361 (7)	156	3.879 (1)	569
Effingham	0.454 (10)	115	1.694 (11)	221
Poorer Predominantly Agriculture				
Hancock	0.512 (12)	52	6.004 (11)	157

NOTE: These counties were selected as representatives of the six county groups identified by principal components analysis--see text for reference.

TABLE 10: DETAILED RESULTS FOR SELECTED ILLINOIS COUNTIES, 5 YEAR PROJECTION, 1967-1972

Employment Sector	Most Urban: Cook Co. Mean % Error	Most Urban: Cook Co. # Employees	Most Suburban: DuPage Co. Mean % Error	Most Suburban: DuPage Co. # Employees	Second Most Urban/Suburban: Will Co. Mean % Error	Will Co. # Employees	Madison Co. Mean % Error	Madison Co. # Employees	Agric. with Large Urban Centers: Champaign Co. Mean % Error	Champaign Co. # Employees	Peoria Co. Mean % Error	Peoria Co. # Employees	Wealthier Predominantely Agriculture: Kendall Co. Mean % Error	Kendall Co. # Employees	Effingham Co. Mean % Error	Effingham Co. # Employees	Poorer Predominantely Agric.: Hancock Mean % Error	Hancock # Employees
10	0.071	1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
11	0.000	0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
12	0.707	82	*	*	*	*	*	*	*	*	0.509	173	*	*	*	*	*	*
13	0.041	51	0.483	14	0.662	45	0.600	3	0.455	20	*	*	*	*	0.455	15	0.667	2
14	0.302	569	0.216	22	0.515	205	0.290	86	0.276	8	0.262	21	0.588	10	0.353	6	1.000	16
19	0.040	8	1.000	9	*	*	0.000	1	*	*	*	*	*	*	*	*	*	*
20	0.021	1591	0.190	550	0.169	163	0.184	282	0.306	646	0.040	134	0.259	21	0.107	8	0.064	16
21	0.003	1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
22	0.365	749	*	*	*	*	*	*	*	*	0.692	27	*	*	*	*	*	8
23	0.025	477	0.390	16	0.470	157	0.130	79	0.090	15	0.228	44	*	*	0.305	74	*	*
24	0.012	66	0.293	41	0.184	18	0.177	26	0.192	15	0.064	25	*	*	0.031	6	0.059	2
25	0.004	58	0.251	297	0.306	52	0.312	5	0.459	61	0.467	43	*	*	0.225	16	1.000	1
26	0.046	1134	0.065	131	0.069	102	0.042	63	*	*	0.001	1	*	*	*	*	*	*
27	0.018	1442	0.390	1088	0.162	157	0.082	41	0.191	188	0.027	49	0.156	5	0.390	187	1.000	76
28	0.110	3821	0.169	137	0.923	3620	1.261	536	0.067	2	0.101	36	1.000	7	1.000	20	*	*
29	0.061	350	0.895	17	0.071	149	0.082	244	*	*	*	*	*	*	0.929	13	1.000	2
30	0.096	2180	0.429	1003	0.149	49	0.365	27	1.000	18	0.603	70	0.360	80	*	*	*	*
31	0.007	42	0.152	5	0.474	18	*	*	*	*	1.000	1	*	*	1.000	133	*	*
32	0.070	904	0.595	349	0.134	159	0.013	32	0.044	6	0.296	58	0.590	23	*	*	0.688	11
33	0.007	360	0.357	500	0.035	68	0.012	157	0.258	98	0.063	405	0.520	117	*	*	*	*
34	0.004	352	0.220	830	0.314	630	0.013	8	0.710	44	0.013	19	0.008	4	1.000	3	1.125	9
35	0.038	2969	0.243	1256	0.003	16	0.030	40	0.213	59	0.012	56	0.313	1362	0.517	15	0.613	19
36	0.011	1299	0.270	658	1.244	331	0.473	224	*	*	0.005	4	0.308	854	0.225	362	0.818	54
37	0.031	860	0.672	252	0.745	143	0.035	38	0.500	39	*	*	0.400	6	*	*	*	*
38	0.015	420	0.741	1072	0.011	1	0.385	5	1.000	11	0.032	1	*	*	*	*	*	*
39	0.009	223	0.521	210	0.006	2	0.037	8	0.082	11	0.004	1	*	*	*	*	0.022	3
Constr.	0.026	2495	0.335	4069	0.046	192	0.284	945	0.048	154	0.100	449	0.094	23	0.129	61	0.015	3
Trans.	0.016	2647	0.285	2131	0.128	1008	0.098	527	0.021	48	0.050	235	0.250	33	0.143	46	0.398	106
Trade	0.025	13900	0.320	13908	0.090	1209	0.035	479	0.002	1024	0.055	1141	0.049	42	0.159	373	0.100	96
Finance	0.021	3477	0.481	3238	0.108	212	0.061	171	0.053	80	0.046	200	0.167	16	0.279	75	0.240	36
Serv.	0.021	9236	0.227	6172	0.150	1503	0.027	299	0.099	893	0.032	535	0.078	38	0.377	765	0.511	539
Other	0.362	963	0.609	330	0.104	11	0.305	29	0.198	17	0.039	12	1.000	12	1.000	20	0.400	2
Total	0.083	1707	0.386	1363	0.280	393	0.184	150	0.264	202	0.172	133	0.361	156	0.454	115	0.512	52
Model	4		5		9		5		5		5		7		10		12	

* No employees in this sector.

D. County Estimates

Counties were grouped on the basis of a principal components analysis of a large number of socio-economic variables.* On this basis, selected counties in each category were chosen to sample the accuracy of the models at this level to determine whether location or other dominating characteristics may have influenced the degree of accuracy obtained. Table 9 provides a summary detailing the best and worst models on the basis of aggregate estimation. The three models, 1, 8 and 11, comprise the complete set of models categorized as providing the worst estimates. Model 5 again dominates, especially in urban counties whereas no one particular model seems to provide consistently reliable estimates of employment in the essentially agriculturally-oriented counties.

When the detailed sectoral results for individual counties are examined in Table 10 for the best model, a reasonable degree of homogeneity of estimate is found within each county. Note that the poorest estimates (those greater than 10 percent) are found for only five of the 31 sectors in Cook County in which employment was observed. In analysing some of the other county estimates, entries of 1.000 signify that either (1) no employment was forecast in that sector for 1972, since none was present in 1967, while, subsequently, new firms have moved into the county between 1967 and 1972 or (2) employment was forecast for that sector on the basis of its presence in 1967 whereas no employment was recorded in 1972, presumably because of firm closure. A convention was adopted to the effect that, in either case, the recorded error would be 100 percent. In one sector in Hancock county, the 100 percent error for sector 20 does in fact represent an overestimate of that order of magnitude rather than signifying the

---

* These groupings of counties were defined as part of a research project supported by the Department of Local Government Affairs, State of Illinois, under the direction of Dr. L.F. Heumann of the Department of Urban and Regional Planning, University of Illinois, Urbana.

TABLE 11

BEST ESTIMATES BY EMPLOYMENT SECTOR, COOK COUNTY

Employment Sector	Single Sector Best Estimate	Model	Estimate Using Model 4 (Best Aggregate Estimate)
10	0.071	All	0.071
11	*	*	*
12	0.707	4, 11	0.707
13	0.004	5	0.041
14	0.171	11	0.302
19	0.040	4	0.040
20	0.015	9	0.021
21	0.003	4, 9	0.003
22	0.105	12	0.365
23	0.004	10	0.025
24	0.012	4	0.012
25	0.000	6	0.004
26	0.003	5	0.046
27	0.009	2, 5	0.018
28	0.037	5	0.110
29	0.001	5	0.061
30	0.029	8	0.096
31	0.003	3, 8	0.007
32	0.043	10	0.070
33	0.007	4	0.007
34	0.001	7	0.004
35	0.004	8	0.038
36	0.011	4, 10	0.011
37	0.015	1	0.031
38	0.005	2	0.015
39	0.003	9	0.009
Construction	0.001	6	0.026
Transportation	0.001	6	0.016
Trade	0.003	12	0.025
Finance	0.003	8	0.021
Services	0.007	7	0.021
Other	0.090	1	0.362
Total	0.045	--	--

existence of one of the two conditions noted above. When sectors from which firms disappeared or into which firms appeared between 1967 and 1972 were removed from the analysis, the aggregate errors dropped as follows:

DuPage	from 0.386 to 0.362
Champaign	from 0.264 to 0.190
Peoria	from 0.172 to 0.131
Kendall	from 0.361 to 0.321
Effingham	from 0.454 to 0.308
Hancock	from 0.512 to 0.451

Obviously, the removal of the sectors for which it is obvious that there have been movements in and out of firms should not be taken to imply that this procedure includes all such firm mobility in all sectors. As was noted earlier, the present data set precludes the making of any statements about firm mobility in sectors in which positive employment was recorded for all years of observation. However, the reduction in error moves some aggregate county estimates towards the levels achieved for the population data analysis. This is one area in which inputs of local expertise from regional planning commissions, local chambers of commerce and the like would make a valuable contribution to the efficiency of estimation.

E. A Hybrid Model

In keeping with the spirit of the population data analysis, an attempt was made to develop a hybrid model. The analysis of the Cook County experience illustrates some of the difficulties. While Model 4 provided the best overall estimate, all of the models were presented at least once in generating the best estimates for individual sectors. This is one area of research in which additional investigation may prove rewarding. What is needed is some systemmatic analysis of the performance of each model cross classified by county type and sector

in the hopes that examination of such a classification might reveal that a combination of one or possibly two models might yield much better estimates at the individual sector level in the counties.

V. SUMMARY AND CONCLUSIONS

The results described in the preceeding section are no better or worse than the sample provided by Greenberg (1972) in his analysis of county employment projections for the New York Metropolitan Region. Generally, it could be stated that of the models tested, Models 4, 5 and 9 generally outperformed the others at a variety of levels--state, multi-county and county. However, there appeared to be little systemmatic variation in predictive ability such that it is felt inappropriate at this time to propose the selection of a preferred model.

Clearly, there are several future directions of inquiry. These are briefly summarized below. One alternative would be to provide a sector by sector analysis of the models in an attempt to explore any explicit spatial dependencies. None of the models considers the influences of adjacent county employment fluctuations in developing forecasts. This spatial dependency problem can now be handled with recently developed statistical techniques (Cliff and Ord, 1973): some of these are discussed in the next section on new model design. However, it may well be that the models tested thus far are too simple to capture enough of the subtletities of employment growth and decline to warrant further analysis. Alternative models do exist, but the additional data requirements are usually prohibitive in terms of eventual implementation. Greenberg (1972, p. 414) commented that

> The added accuracy increment (from the assembly of additional data) was not sufficient to justify the added expense.

In the next section, some alternative models which have few additional data requirements (and these are, for the most part, readily available) are described in terms of their potential application to the Illinois data. A third alternative would be to concentrate on a small number of industries deemed to be existing or future potential polluters and develop models to forecast their employment levels. Consideration would also be given to those industries which are

major suppliers to the problem industries or major consumers of their output. In this fashion, the analysis could be conducted at the three and four digit SIC level utilizing the available data from the County Business Patterns.

A fourth alternative would be to concentrate activities on forecasting employment and other related activities within an SMSA or another set of county units. The Appendix to this Report details some of the models that are currently available to facilitate this type of analysis. They do, however, have one problem in common: they rely heavily on survey data and hence require the commitment of rather large sums of research dollars.

In terms of ease of implementation, alternatives one and two above could be handled with the additions of a small set of data. Alternative three would require the assembly of a new data file from previous volumes of the County Business Patterns. In addition, a certain amount of time would be required to "clean up" the data--to estimate employment in industries not disclosed at the three and four digit level. The last alternative would require varying amounts of time, depending upon the model chosen--a survey based economic-ecologic input-output model for an SMSA would probably represent a two or three year undertaking, whereas a nonsurvey economic base model could be developed in a much shorter period of time.

In the next section, some guidance in the context of alternatives one and two is provided.

VI. NEW MODEL DESIGN

The models utilized in stepping down county population and multi-county employment projections have two general drawbacks: (1) with the exception of one employment model, they all ignore the interdependence between population and employment levels and (2) spatial interaction and interdependence is ignored. All of these issues will be discussed within the context of alternative models with the caveat that there is no guarantee that more sophistication in model design will result in more accurate predictive capability. It was pointed out that some of the simpler population models achieved greater accuracy than other models which, a priori, one would have expected to have achieved more accurate results as a result of their supposed greater realism.

INTERDEPENDENCE OF POPULATION AND EMPLOYMENT [CZAMANSKI'S MODEL]

In the design of the Illinois Bureau of the Budget (IBOB) employment and projection model, considerable attention is focused on the role of population--through its impact on participation rates and migration. At the township level, it may make little sense to justify the expense of preparing a data base that would enable one to replicate an IBOB-type model. However, at the county level, this may not only be possible and feasible but rather important as a further check on the reliability of both population and employment forecasts.

Czamanski (1964, 1965) has proposed a model which explicitly deals with the joint forecasts of population and employment for urban areas. In notation, the system may be expressed as follows:

$$P = f_1 (E) \tag{1}$$

$$E = E_g + E_c + E_u \tag{2}$$

$$E_c = f_2 (E_g) \quad (3)$$

$$E_u = f_3 (P) \quad (4)$$

where: P,E are population and employment respectively in the city

E_g is employment in industries whose main locational factors are independent of considerations of localization or urbanization,

E_c is employment in industries whose main locational factor is the presence of other industries to whom they supply or from whom they draw goods and services,

E_u is employment in industries whose main locational factor depends on the existence of the city.

Equation (1) relates population to total employment: the latter may be decomposed into three categories as is shown in the identity equation (2). Equation (3) represents the linkages between two of the industrial categories while equation (4) completes the system by relating the third employment category to total population. Obviously, the theoretical formulation requires careful translation into empirical implementation especially when the application will be in the context of county units rather than (exclusively) urban areas. In the case of the SMSA counties, the empirical formulation could follow that suggested by Czamanski (1965): the issues associated with application in non-SMSA counties will be discussed below.

For the application of the model to the Baltimore SMSA, Czamanski (1965) obtained the three employment categories as follows. First, the urban oriented industries were designated through regressing population and employment in each industry:

$$E_i (t) = a + b P(t-1) \qquad i = 1...n$$

n = total number industries

It was found that the highest correlations were obtained with a one year lagged equation system. Those industries with the highest correlations were assumed to be urban oriented, since their growth was associated with population change. There remained the problem of dividing the remaining industries into the other two categories, E_c and E_g. The former category can be thought of as complementary industries in the sense that they have strong input-output linkages with a small set of other industries within the urban area. Czamanski claimed that, in the absence of local input-output data, the complementary industries could be identified by assuming (1) that their increases and decreases in output would be associated with increases and decreases in output in the industries to which they are considered complementary (perhaps adjusting for a suitable time lag) and (2) the industry should be small in the sense that it would employ less than 0.5 million in the nation as a whole. The rationale for the last category was never made clear. To implement the first identification procedure, Czamanski again used correlation analysis, experimenting with various lagged and unlagged relationships. However, he was only able to identify one complementary industry! The remaining industries were categorized as geographically oriented (i.e. E_g). The system of equations was as follows:

$$P(t) = a_1 + b_1 E(t-2) \tag{5}$$

$$E(t-2) = E_g(t-2) + E_u(t-2) + E_{PMI}(t-2) \tag{6}$$

$$E_u(t) = a_2 + b_2 P(t-1) \tag{7}$$

$$E_{PMI}(t-2) = a_3 + b_3 E_{BF}(t-2) \tag{8}$$

where: P, E, E_g, E_u were defined as before

E_{PMI} was employment in Primary Metals Industries, the only complementary category,

E_{BF} employment in Blast Furnaces.

E_{BF} and E_g were assumed exogenous variables: the remainder, endogenous. The reduced form yielded:

$$P = A\,(1 + C + C^2 + C^3 \ldots C^{n-1}) + b_1\,(E_g(t-2) + CE_g(t-5) + C^2E_g(t-8) + \ldots C^{n-1}E_g(t-3n+1)) + B\,(E_{BF}(t-2) + CE_{BF}(t-5) + C^2E_{BF}(t-8) \ldots + C^{n-1}E_{BF}(t-3n+1)) + C^nP(t-3n) \quad (9)$$

where: $A = a_1 + b_1\,(a_2 + a_3)$

$B = b_1b_3$

$C = b_1b_2$

The model yields two multipliers, A and B, relating forecast population to past employment and one multiplier, C, relating forecast population to present population.

Czamanski found that the geographically oriented industries exercised the greater influence in population size than the other multipliers (i.e., $B > A > C$). Czamanski forecast 1964 through 1980 population for Baltimore SMSA using his model: the actual population was 2.030 m, the estimate 2.071 m, an error of approximately 2%. Czamanski used various forms of shift analysis to project E_g and E_{BF}.

Application to the Current Research Project

Czamanski's model may provide more reasonable estimates of population and employment within the SMSA counties. However, the author did note in an earlier paper that the classification of industrial employment into the three components of the model varied according to city size. Hence, the model would have to be calibrated separately for each of the SMSA's. For the non-SMSA counties, it makes little sense to assume that

their population and employment totals will be independent of those in the SMSA's. Thus, an alternative model may be required for these counties, one which specifies the employment and/or population totals for the nearest SMSA as exogenous variables. In this case, estimates will be required for the St. Louis, Mo. SMSA for input into the model for counties in the St. Louis field of influence. Furthermore, population projections have already been made for Illinois counties and hence population cannot be included in the model as an endogenous variable. Application of the model to Illinois counties would have to take on a slightly different character whose reduced form might specify multi-county or SMSA estimates of certain employment categories together with total population as the exogenous variables.

In summary, the notion of joint estimation of population and employment is attractive in situations in which only limited forecast data are available. Inasmuch as estimates of county population and multi-county employment levels are already available, the model, as Czamanski specified it, would be reduced in form to a three equation model in which employment is subdivided into the following components:

(1) employment which is sensitive to changes in population (Czamanski's E_u),

(2) employment which is functionally related to other types of employment change (Czamanski's E_c),

(3) employment which bears a strong relationship to movements in that category at the multi-county level (Czamanski's E_g).

In addition, the independent variables would be expanded to include activity in the nearest SMSA in the equation system for non-SMSA counties.

SPATIAL INTERACTION AND SPATIAL INTERDEPENDENCE

In the preceeding discussion, it was hinted that employment and population changes in other parts of the state may influence the levels in another county. The adjustments of township population estimates to conform with county totals, and county employment estimates to conform with multi-county totals do provide a measure of de facto recognition of spatial influences. However, it still assumes that the population estimate for township j ($j \varepsilon J$, where J is the county) is independent of that for township k ($k \varepsilon J$) and similarly for employment estimates at the county and multi-county levels. Cliff and Ord (1969, 1971, 1972, 1973) have been undertaking some pioneering research into the problem of spatial autocorrelation. Some aspects of their research are presented below, especially those ideas which may have particular relevance for future model design.

Let the space, i.e., the state, be divided into n non-overlapping counties. Let the observed value of a variable X in a typical county be x_i.

No spatial autocorrelation exists if the x_i and x_j for any pair of counties i and j are uncorrelated. Spatial autocorrelation exists if the pairwise drawings of counties are not all uncorrelated. The problem with spatial autocorrelation is that spatial dependence in a county system may extend in all directions: in time series data, serial correlation relates only to the influence of values of the variate x_i at t-1, t-2, t-3, ... on the values of x_i at t. Hence, even first-order spatial autocorrelation may involve consideration of the influence of k ($k \varepsilon n$) counties while higher order spatial autocorrelations may involve the

complete county system. The latter is unlikely, since there will probably be a distance decay effect in existence.

Let $||\delta_{ij}||$ be a connection matrix:

$$\delta_{ij} = \begin{cases} 0 \text{ if } i,j \text{ counties are not joined} \\ 1 \text{ if } i,j \text{ counties are joined} \end{cases}$$

Moran's coefficient is given by:

$$I = \frac{n}{2A} \frac{\sum_i \sum_j \delta_{ij} z_i z_j}{\sum_i z_i^2}$$

where: $z_i = x_i - \bar{x}$

A is the total number of joins in the county system

$= 1/2 \sum_i L_i$ where L_i is no. counties jointed to the ith county

(factor 1/2 eliminates duplication of counting both ij and ji joins).

Geary's statistic c is given by:

$$c = \frac{n-1 \sum_i \sum_j \delta_{ij} (x_i - x_j)^2}{4A \sum_i z_i^2}$$

In both cases, the numerator is a measure of covariance among the variable x's and the denominator a measure of variance. The moments of I and c may be evaluated under assumptions of normality or randomization (see Cliff and Ord 1973, pp. 8-9).

The I and c statistics make no reflection of the size and shape of counties in the system and the relative strengths of linkages between them (i.e., how many road/rail/air links). Thus I and c values could be obtained for topologically identical sets of counties which in real space had very different spatial properties.

Dacey suggested a new I, I':

$$I' = \frac{n}{2A} \frac{\sum_i \sum_j \delta_{ij} \alpha_i \beta_{i(j)} z_i z_j}{\sum_i \alpha_i z_i^2}$$

where: $\alpha_i = \frac{a_i}{\sum_i a_i}$ $\qquad$ $\beta_{i(j)} = \frac{b_{ij}}{\sum_{j \epsilon J} b_{ij}}$

a_i is the area of ith county,

b_{ij} is length of common boundary between i and j counties, and

$j \epsilon J$ provides summation over all j counties joined to i.

All measures discussed thus far take into account the influence of spatially contiguous counties only: there is no distance decay effect, i.e., measuring the impact that county k which is joined to county j has on county i.

Cliff and Ord suggest replacing the binary weights δ_{ij} with a generalized weighting matrix $W = ||w_{ij}||$.

Thus $I = \frac{n \sum_i \sum_j w_{ij} z_i z_j}{w \sum_i z_i^2}$ $\qquad$ where: $w = \sum_i \sum_j w_{ij}$

$$c = \frac{n-1}{2w} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i z_i^2}$$

Advantages (and disadvantages) of this weighting method are that it allows the choice of a set of weights a priori. Thus, hypotheses about the influence of the degree of contact between counties may be investigated. For example, it may be assumed that the amount of interaction between any two counties may be a function of the distance between their major geographic or population centers, the structure of the transportation system,

or the length of common boundary between the two countries. One example given by Cliff and Ord (1973) suggests that the weights take on the following form:

$$w_{ij} = d_{ij}^{-a} (\beta_{i(j)})^{b}$$

where: d_{ij} is the distance between the centers of counties i and j and

$\beta_{i(j)}$ is the proportion of the perimeter of county i which is in contact with j.

Hence, $\sum_{j \epsilon J} \beta_{i(j)} = 1$ where J is the set of counties contiguous to county i.

These tests can thus be applied to a set of counties to measure the possible influence of any county estimate of population or employment on the estimates for another county. Of particular interest in terms of further model design, is the application of these tests to the analysis of residuals. Thus, in addition to usual tests of serial correlation, new tests may be applied to the data to determine whether or not they are spatially autocorrelated. These tests would enable one to determine whether autocorrelation exists because of misspecification of the relationship between the variables in say a regression equation.

The spatial autocorrelation tests together with mapped output of the residuals would provide useful insights into the possible causes of over and underprediction in the various models of employment and population projection. With a number of automated mapping routines currently available, this type of analysis is now eminently feasible and should serve to provide an important additional dimension to the success of various models.

INDUSTRIAL INTERDEPENDENCE

In addition to the absence of concern with the spatial aspects of growth, the employment models used in this report do not utilize currently available information about the structure of industry. This aspect has been explored in greater detail by Chalmers and Beckhelm (1974) in a model they designed linking shift and share analysis with industrial location theory. In a sense, it represents an attempt to specify more vigorously the idea incorporated in Czamanski's (1964) model in which a set of industries, entitled complementary, was defined. Chalmers and Beckhelm (1974) sought to challenge Brown's (1969) assertion that the competitive component of industrial growth (see below) is not associated with variables related to industrial location theory.

Refering back to model 8 of the Employment Projections, it will be recalled that employment growth in any industrial category was assumed to be a function of two major components which are usually referred to as the shift (or competitive component) and the share component. In notation recall that

$$E_{t+5}^{ijk*} = [r^{jk*} + (r^{ijk} - r^{jk})]\, E_t^{ijk}$$

where E_{t+5}^{ijk*}, E_t^{ijk} are the forecast and observed employment in class k in county i ($i \varepsilon j$, the multi-county) at time t+5 and t respectively,

r^{jk}, r^{jk*} are the rates of growth of employment class k in the jth multi-county between t-5 and t and t and t+5 respectively,

r^{ijk} is the rate of growth of employment class k in county k between t-5 and t.

Thus, r^{ik*} represents the share component and $(r^{ijk} - r^{jk})$ the shift or competitive component. Chalmers and Beckhelm (1974) attempted to explain the latter component through the use of a set of explanatory variables which are described in greater detail below.

The objective, then was to relate the competitive component variables assumed to be important. The variables were as follows:

Consumer Market Potential ($CMP_t(t)$) and Relative Consumer Market Potential ($RCMP_r(t)$)

For region r:

$$CMP_r(t) = \sum_{i=1}^{m} P_i^t/\delta_{ri}^{\beta}$$

where: $P_i(t)$ is population in region i for time t

δ_{ri}^{β} is the distance from region r to region i

$\beta = 1.5$ $(r \neq i)$

$\delta = 1.0$ for $r = i$

$$RCMP_r(t) = CMP_r(t)/(\sum_{i=1}^{m} CMP_i(t)/m)$$

Intermediate Market Potential ($IMP_{kr}(t)$) and Relative Intermediate Market Potential ($RIMP_{kr}$)

For industry k in region r:

$$IMP_{kr}(t) = \sum_{g=1}^{n} (\sum_{i=1}^{m} E_{gi}^{\alpha}(t)/\delta_{ki}^{\beta}) \ (S_{kg}^{63}/E_{g.}(t))^{\alpha}$$

where: n is the number of industries,

m is the number of regions

E_{gi} is total manufacturing employment in industry g in region i,

$E_{g.}$ is total industry employment in industry g,

S^{63}_{kg} is total sales for industry k to industry g obtained from the 1963 Input-Output Table for the U.S.,

δ_{ri} is the distance from r to i,

α and β are parameters.

$$RIMP_{ki}(t) = IMP_{ki}(t) / \left(\sum_{i=1}^{m} IMP_{ki}(t)/m \right)$$

Supply Potential ($SP_{kr}(t)$) and Relative Supply Potential ($RSP_{kr}(t)$)

For industry k in region i:

$$SP_{kr}(t) = \sum_{k=1}^{m} E_{ki}^{\alpha} / \delta_{ri}^{\beta}$$

E_{ki} is total manufacturing employment in industry k in region i,

δ_{ri} is as before.

$$RSP_{kr}(t) = SP_{kr}(t) / \left(\sum_{i=1}^{m} SP_{ki}(t)/m \right)$$

Wages ($W_{ki}(t)$) and Relative Wages ($RW_{ki}(t)$)

For industry i in region j:

$$W_{ki}(t) = PR_{ki}(t)/MH_{ki}(t)$$

where: PR_{ki} is total manufacturing payroll for industry k in region i

MH_{ki} is total manufacturing man hours in industry k in region i.

$$RW_{ki}(t) = W_{ki}(t)/WA(t)$$

where: $WA(t)$ is the average wage rate for manufacturing in year t.

$$= \frac{\text{Total US Manufacturing Payroll}}{\text{Total US Manufacturing Man Hours}}$$

Supplier Availability ($SAP1_{kr}(t)$, $SAP2_{kr}(t)$) and Relative Supplier Availability ($RSAP1_{kr}(t)$, $RSAP2_{kr}(t)$)

For industry k in region r:

$$SAP1_{kr}(t) = \sum_{i=1}^{m} E_{1i}^{\alpha} / \delta_{ri}^{\beta}$$

$$SAP2_{kr}(t) = \sum_{j=1}^{k} E_{2i}^{\alpha} / \delta_{ri}^{\beta}$$

where: E_{1i} and E_{2i} represent the two top industries from which industry k obtains intermediate goods. This information is derived from the 1963 Input-Output Tables for the U.S.

$$RSAP1_{kr}(t) = SAP1_{kr}(t) / \left(\sum_{i=1}^{m} SAP1_{ki}(t)/k \right)$$

Similarly for $RSAP2_{kr}(t)$.

The model, relating the differential shift to the independent variables described above*, was run for 23 two and three digit industries for various SMSA's using the years 1963 and 1967. (The number of SMSA's varied according to disclosure of information: the range was 13 for periodicals to 67 for apparel.) The results indicated that the competitive component could be explained by measures of consumer and market potential and supply. It was also found that changes in the values of these variables were often more important than their relative values--i.e., ΔCMP_r rather than $RCMP_r$.

In the context of the Air Quality Maintenance Program, the Chalmers and Beckhelm model may be useful in capturing the interdependencies of growth among the set of industries most closely allied with major air

---

* In addition, change variables, i.e., ΔCMP_r were also calculated.

polluters, especially in SMSA counties. It remains to be seen just how well it would predict employment in non-SMSA counties and in employment classes not characterized by a great deal of interindustry dependence.

SUMMARY

The extensions to the models used in the current analysis should not be considered all inclusive. Many other directions for future research are possible: the suggestions made here incorporate ideas which would utilize the existing data base wherever possible, requiring only marginal effort in additional data assembly. For example, annual population estimates for counties would be required to implement the Czamanski model, national input-output tables for 1963 or 1967 would be required for the Chalmers and Beckhelm model and certain data on spatial contiguity would have to be assembled and calibrated to allow testing for the presence of spatial autocorrelation.*

Hence, additional research can be directed almostly entirely to model design and testing rather than data assembly. Given the relatively poor estimates provided by the employment models, it would make sense to extend the analysis to consider alternative formulations. However, the relative volatility of county employment cautions one to repeat the statement made at the introduction, namely that greater sophistication in model design is no guarantee of greater accuracy of forecasting county employment levels. Minor fluctuations in one firm's employment can often

---

* Two sets of data have already been keypunched:
(1) a connection matrix $||\delta_{ij}||$ where $\delta_{ij} = 0$ if i,j counties are not joined and 1 if i,j counties are joined and
(2) a common boundary matrix $||b_{ij}||$ where b_{ij} is the length of common boundary between i and j counties.

create substantial errors in the forecasts made by even the most sophisticated model. Undoubtedly, there are diminishing returns at some point in model design but with the best mean percentage error in excess of 50%, it would not appear unreasonable to suggest improvements at reasonable cost could still be made.

Additional models, suitable for analysing population and employment (among other parameters), within a single region (for example, a multi-county or an SMSA) are described in some detail in an appendix to this report. These models presuppose a commitment to analysis at a single region level rather than for a set of regions within a state.

REFERENCES

Argonne National Laboratory (1973). State of Illinois Water Management Program, Appendix A The Socio-Economic/Land Use Model, SELUM, Center for Environment Studies, Argonne National Laboratory, Argonne, Illinois.

Ashby, L. D. (1970). "Changes in Regional Industrial Structure: A Comment," Urban Studies, 7, pp. 298-304.

Brown, H. J. (1969). "Shift and Share Projections of Regional Economic Growth," Journal of Regional Science, 9, pp. 1-18.

Chalmers, J. A. (1971). "Measuring Changes in Regional Industrial Structure: A Comment on Stillwell and Ashby," Urban Studies, 8, pp. 289-292.

Chalmers, J. A., and Beckhelm, T. (1974). Shift and Share and the Theory of Industrial Location, Faculty Working Paper 74-18, Department of Economics, Arizona State University, Tempe.

Cliff, A. D., and Ord, J. K. (1969). "The Problem of Spatial Autocorrelation," in A. J. Scott (ed.) Studies in Regional Science (Pion, London), pp. 25-55.

Cliff, A. D., and Ord, J. K. (1971). "Evaluating the Percentage Points of a Spatial Autocorrelation Coefficient," Geographical Analysis, 3, pp. 51-62.

Cliff, A. D., and Ord, J. K. (1972). "Testing for Spatial Autocorrelation Among Regression Residuals," Geographical Analysis, 4, pp. 267-284.

Cliff, A. D., and Ord, J. K. (1973). Spatial Autocorrelation (Pion, London).

Czamanski, S. (1964). "A Model of Urban Growth," Papers Regional Science Association, 13, pp. 177-200.

Czamanski, S. (1965). "A Method of Forecasting Metropolitan Growth by Means of Distributed Lags Analysis," Journal of Regional Science, 6, pp. 35-49.

Greenberg, M. R. (1972). "A Test of Alternative Models for Projecting County Industrial Production at the 2, 3 and 4-digit Standard Industrial Code Levels," Regional and Urban Economics, 1, pp. 397-417.

Hellman, D. A. (1974). "A Model of Regional Export Activity," Growth and Change, 5, pp. 12-17.

Hellman, D. A., and Marcus, M. (1970a). A Critical Analysis of Employment Projection Methods: A Test Case of New Jersey, Part II of Urbanization and its Effect on Water Resources, Water Resources Research Institute, Rutgers University, New Brunswick, New Jersey.

Hellman, D. A., and Marcus, M. (1970b). "The Temporal Stability of Localization Quotients: An Empirical Evaluation," Economics and Business Bulletin, 22, pp. 1-17.

Houston, D. (1967). "Shift and Share Analysis: A Critique," Southern Economic Journal, 32, pp. 577-581.

James, F. J., and Jughes, J. (1973). "A Test of Shift and Share Analysis as a Predictive Device," Journal of Regional Science, 13, pp. 223-231.

OBERS (1972). OBERS Projections of Regional Economic Activity in the US, Volume I Concepts, Methodology and Summary Data, US Water Resources Council, Washington, D.C.

Stillwell, F. J. (1969). "Regional Growth and Structural Adaption," Urban Studies, 6, pp. 162-178.

APPENDIX I

MORE SOPHISTICATED MODELS FOR SINGLE AREA ANALYSIS

I. Introduction

Development of effective control strategies for the Air Quality Maintenance Planning Program requires input from a number of diverse fields. Obviously, scientific and engineering expertise are necessary for the identification of the sources of air pollutant emissions, the determination of their amounts and potential harm and the development of control technology to eliminate or reduce these emissions. This information alone, however, is not sufficient to overcome our present and future air quality problems.

As described in volume 4 of the Guidelines for Air Quality Maintenance Planning and Analysis, consideration must be given to inputs from the fields of Economics and Planning. Detailed projections of the expected increases in the various sources of emissions must be made. In a very simplified sense, this means projections of both population and economic activity (employment). Planning to meet future emission levels can only be effective with a knowledge of what kinds and amounts of pollution sources are to be expected.

Several approaches to making the above mentioned projections are contained within this report, included are:

(1) Economic Base Analysis

(2) Input - Output Analysis

(3) Hybrid Input - Output Models

(4) Economic - Ecologic Linkages Models

(5) Econometric Models

These techniques, varying in complexity and operational costs, provide

a wide range of alternatives to making detailed population and employment projections. Each method is subject to limiting assumptions, and none can, (nor can any urban structure model) claim the distinction of precisely reflecting future conditions. However, they are capable of providing varying degrees of detailed information on the existing economic structure of an area, and, based upon this structure, generating estimates of the likely economic growth of that area.

While population and employment projections are necessary and useful, they in and of themselves, are not sufficient to meet the needs of the air quality maintenance program. Using these projections as control totals, estimates must be made of the spatial location of this anticipated growth. The final section of this report reviews the problems of allocating future land uses to specific areas within a region, and also includes summaries of several relatively inexpensive and operational land use allocation models. The output of these models can provide useful information for making estimates of the location of expected future sources of pollutant emissions.

II. Economic Base Analysis

II. 1. General Description and Theory

In the context of economic base theory, economic activities of an area are divided into two broad types: the basic activities which produce goods and services for export to firms and individuals outside a defined economic area, and the service or non-basic activities whose output of goods and services is consumed within the economic area. Thus, it seeks to distinguish between productive activities which bring new money into the community (basic activities) and those activities which merely recirculate money already existing within the community (service activities). Economic Base Theory contends that the basic activities are the key to the economic growth of the community, and that growth in basic output results in growth in the service activities and thus growth in the total economy. Obviously, much of the validity of the base concept lies in being able to classify the area's economic activities as being either basic or non-basic. In notation, let

Y_T be total income generated by all activity in the area

Y_B be that portion generated by exports (i.e. basic activity)

and Y_{NB} be the non-basic portion of income.

By definition

$$Y_T = Y_B + Y_{NB}$$

If $r = Y_{NB}/Y_T$ (the proportion of total income that is non-basic income), then

$$Y_T = Y_B + rY_T$$

Rearranging and factoring

$$Y_T = \frac{1}{(1 - r)} Y_B$$

The economic base multiplier is $1/(1 - r)$: this multiplier is then used in conjunction with estimates of changes in basic output (employment) to project various growth components for the community.

II. 2. Uses of a Base Study

II. 2. A. Description of Existing Conditions

The types of general information generated by an economic base study are of interest to planners, public decision-makers, businessmen and development groups. A few of the many benefits gained through such a study are: (1) an identification of current sources of income and employment; (2) an identification of weaknesses in the local economy, e.g., how much basic employment is tied to one industry; and, (3) an identification of the types of industries which should be encouraged to expand or locate within the area. Furthermore, periodic base studies can aid in evaluating a community's progress toward achieving certain public goals such as increasing total employment, and increasing per capita income. A base study also serves to identify the industrial mix of the export sectors. If this mix is sensitive to cyclical fluctuations, an appropriate policy may be to encourage industries of a more stable nature. Finally the study can provide valuable information to business firms both within and outside the area by identifying local market potential, local sources of demand for their products, and the relationship of a firm's inputs and outputs to other industries within the local economy.

II. 2. B. Projection

The key factor in projecting community growth, as mentioned earlier, is the basic or export market; specifically, how it will grow or decline in the coming years. Population and related considerations are projected by constructing a series of ratios using the base multiplier and the estimated change in basic output. To demonstrate this procedure assume projections of (1) total change in population and (2) total change in single family housing units are desired. Such projections are obtained as follows:

$$\frac{1}{(1 - r)} \Delta Y_B = \Delta Y_T$$

$$\Delta Y_T \cdot (P/Y_T) = \Delta P \quad : \quad \left(\begin{matrix}\text{Change in Total} \\ \text{Employment}\end{matrix}\right) \left(\frac{\text{Population}}{\text{Total Employment}}\right) = \begin{matrix}\text{change in} \\ \text{total} \\ \text{population}\end{matrix}$$

$$\frac{1}{(1 - r)} \cdot \Delta Y_B \cdot (P/Y_T) \cdot (HU/P) \cdot (SFHU/HU) = \Delta SFHU$$

where Δ denotes "change in"

P population

HU housing units

SFHU single family housing units

Obviously this procedure could be continued to project a number of items such as land area, sales, commercial floor space, etc. The validity of this approach is limited in that it makes the following assumptions: (1) For urban population growth to occur, it is necessary that the expanded labor requirements not be furnished by those presently unemployed, i.e., some in-migration is necessary; (2) The ratios remain constant over time; and (3) The average figures with which the ratios are calculated accurately describe present and future conditions, e.g., average incomes accurately represent the

existing and future transactions within the community.

II. 3. Indirect Methods for Determining the Base Ratio

Although the following methods are described using employment as the unit of measurement, several other measurements may be used. The others most often identified are as follows: (1) sales; (2) value added; (3) income (as in the discussion above); (4) payroll; and, (5) total production. Each has certain advantages and disadvantages, and the reader is referred to any of the various texts on economic base analysis. (Pfouts, 1970, Tiebout, 1962).

II. 3. A. Assignment

The simplest method of identifying basic and non-basic sectors is merely to assign the local firms to either category using one's own knowledge of the activities within the area. The usual assumption is that all manufacturing and agriculture is export and the rest is service. Obviously such an approach is largely subjective, and therefore subject to enormous error. For example, a significant portion of manufacturing's output may be locally oriented. On the other hand many service firms, e.g., insurance home offices, may be largely non-locally oriented. The method's simplicity makes it attractive, but it is a highly unreliable approach and as such, is not recommended.

II. 3. B. Location Quotients

The location quotient is based on the assumption that a region's consumption of a good is proportional to its share of employment compared to some base, usually the nation, (although the state, county,

etc. may be used). It is calculated by the following ratio:

$$L.\ Q._i = \frac{\%\ \text{of region's employment in industry } i}{\%\ \text{of nation's employment in industry } i}$$

A ratio of less than or equal to 1 assumes that all production in i is consumed within the region and thus all employment is non-basic. A ratio greater than 1 assumes that i's production is greater than the region's demand for that good and therefore is exported. The amount being allocated to export being expressed as (1/L.Q.) times the employment in industry i. While this approach is useful and often employed, certain problems are inherent in the method and need to be stated: (1) The approach assumes that the regional consumption is identical with the average consumption of the nation, i.e. it ignores regional variations, preferences and tastes; (2) It assumes an average level of productivity/unit of employment in each region; (3) When the L.Q. = 1, it assumes the region consumes all that it produces, ignoring the possibility of cross hauling; (4) The ratio varies with the level of industrial aggregation; (5) It understates exports, thus overestimates the multiplier. It does however, have the advantage that it identifies indirect exports, i.e., output flows from one local firm (service) to another, which are used as an input and then exported (basic).

II. 3. C. Minimum Requirements (Ullman and Dacey, 1960)

The minimum requirements approach begins by calculating the percent of the total labor force employed in each industry for several communities similar to the one under study. It then assumes that the minimum percentage in a given industry in any community is all that is needed to meet local demand. Therefore, all employment above

this amount is considered export employment. Repeating this procedure for each industry yields total export employment for the study community. Several problems exist with such an approach: (1) The community with the minimum percent may be importing, i.e., the minimum isn't enough to satisfy local demand; (2) The study community may have a local demand greater than that of the "minimum" community; (3) Differences in tastes - there may be little demand for the good in the minimum requirements area; and, (4) It assumes local needs are met by local production - no import substitution.

II. 3. D. Exogenous Regression (Mathur and Rosen, 1974)

This technique assumes that the level of exports within a region is in response to economic activity in the "world." Statistically this can be described and analyzed by the following relationship:

$$E_r^i = A + B\,E_w^i$$

where E_r^i = Region's employment in industry i

E_w^i = "World" employment in industry i

A,B = Constants determined through regression analysis

Total employment is divided into non-basic and basic by the following relationship:

$$1 = \frac{A}{E_r^i} + B\,\frac{E_w^i}{E_r^i}$$

The first term to the right of the equal sign represents non-basic employment, the second term representing basic. The problems of such an approach are: (1) Determining what constitutes the "world"; (2) The correlation between regional employment and world employment

depends upon the regional trade patterns within subsets of the "world"; and, (3) It assigns the entire portion of regional employment which is correlated with world employment to basic. However, by economic base theory:

non-basic = f(basic)

by Mathur and Rosin:

basic = f("world")

therefore:

non-basic = f("world")

Thus since some non-basic employment is correlated with world employment, assignment of all employment that is correlated with world employment to basic, results in an overestimate of basic employment and thus an underestimate of the multiplier. Because the L.Q. overestimates the multiplier, the two could be used in conjunction to establish upper and lower limits for the base multiplier.

II. 4. Direct Methods for Determining the Base Ratio

II. 4. A. Survey

Determination of what proportion of employment is engaged in activities which export is possible by means of a survey of local economic activities. Usually employing sampling techniques, such a survey attempts to differentiate between the percentage of sales which were local and those which were non-local. These sales percentages are then applied to total employment figures to obtain estimates of basic and non-basic employment. However such a procedure is prone to the usual problems of sample validity, sector

aggregation, product mix, and indirect exports. In addition, due to disclosure rules and firm idiosyncrasies, the requested data are often unavailable. Finally, such a procedure can require considerable expenditure in time and money.

II. 4. B. Iterative

The iterative process (Tiebout, 1962) attempts to account for the problem of indirect exports. The technique begins by allocating the output of each industry to three components: exports, local consumption and other local firms. The output from industry i to local industry j is then allocated to export, local consumption and other local firms based upon j's total output to these three categories. This still results in some of i's output going to local firms. This procedure is then continually iterated until the local firm allocations are negligible. By repeating this process for each i within the region, all indirect exports can be traced and thus a more accurate estimate of total basic employment is possible.

II. 5. The Economic Base Multiplier Model

As stated earlier increases in the employment and income of a region are assumed to be dependent only on the level of basic activity in the region. Therefore to make projections with the base multiplier, one must be able to project the future demand of the export sectors.

II. 5. A. Forecasting the Export Sector

The most frequent approach would be to simply investigate, using published data, the future growth prospects of the region's export industries. A second approach could be to ascertain the community's

resources and its locational advantages and disadvantages. From this study it may be possible to then identify certain industries which could feasibly locate in the region. The third approach simply assumes that the study region will continue to maintain its present share of national employment. Using estimates of this share and national projections, the region's future export size is then estimated. One final approach is based upon the knowledge and judgement of local business leaders. It is presumed that top executives, bankers, etc. are concerned about the growth of their firm, and therefore, are knowledgeable about future levels of activity.

II. 6. An Alternative Multiplier

This multiplier recognizes that the income flowing into a region is dependent upon a number of exogenous factors. Total production (Y) can be allocated to three sectors: Endogenous consumption (C) and two exogenous components, local investment (I) and government spending (G).

$$Y = C + I + G$$

Assuming consumption is a function of production, $C = cY$, where c = the marginal propensity to consume, i.e., the porportion of total earnings respent, and substituting this relationship into the original equation yields:

$$Y = cY + I + G$$

$$Y - cY = I + G$$

$$Y = \frac{1}{1 - c} (I + G)$$

Thus this multiplier approach views growth in production as being

related to factors both within and outside the community. Changes in production are found by substituting changes in any or all of the factors in the above equation. The advantage of such an approach is that it permits greater disaggregation of production, thus enabling one to better understand the complex factors underlying regional growth. The preference of such disaggregation over the base multiplier, which relates production growth solely to exports, should be obvious.

This multiplier concept may be carried one step further to include consideration of exports (X) and imports (M).

$$Y = C + I + G + X - M$$

Imports, as with consumption, is assumed to be a function of production, $M = mY$, where m = the marginal propensity to import, i.e., that proportion of total imcome which is spent outside of the region for imports. Substitution yields:

$$Y = cY + I + G + X - mY$$

$$Y(1 - c + m) = I + G + X$$

$$Y = \frac{1}{(1 - c + m)} (I + G + X)$$

Again, local production is related to several exogenous factors which are not considered by the base multiplier.

Despite the advantages of such a multiplier approach, there are several disadvantages. First of all, the marginal propensity to consume is extremely difficult to quantify. Indeed, its value is believed to change with different income classes, residents versus immigrants, and regions. Similar statements could be made for the

marginal propensity to import. The obvious limitation of the model is that one must accurately predict the changes in each variable on the right side of the equation in order to project changes in production, or income. However, its identification of the various components of growth can aid the analyst in his study of a regional economy.

II. 7. Criticisms of Economic Base Analysis

There have been numerous past debates on the importance of the service or non-basic activities in influencing regional growth. Advacates of economic base theory argue that it is the basic activities which bring dollars into the area and thus are the key to the area's growth. However, others have pointed out that if a region does not have a well established service sector, many items will have to be imported, thus making location within the region economically unfeasible.[1]

Numerous criticisms of the base approach have been advanced through the past years: the following is a summary of several of the more significant ones. First, there is the problem of defining the study area - often dictated by available data. The size of the area critically affects what production is considered basic and non-basic. For example a good which flows from the CBD to the county would be basic if the city limits define the study area, and would be service if study area was synonymous with the county. Second,

[1]For the implications and details of this type of debate, see C. Tiebout, "Exports and Regional Economic Growth", Journal of Political Economy, April 1956, including "Reply" by North and "Rejoinder" by Tiebout reprinted in J. Friedman and W. Alonso, Regional Development and Planning (MIT Press 1964).

there is the problem of the unit of measurement, as mentioned earlier. The third and perhaps major criticism is the problem of identifying the basic and non-basic sectors. This is obviously complicated by the existence of product mixes, indirect exports, and industrial aggregation. Fourth, the multiplier approach predicts future events based upon past relationships. Such a criticism is common in most if not all projection techniques, but nonetheless should be mentioned. Fifth, the ratio of non-basic/total activity at any one time (which is subsequently used for projection purposes), is most likely to prove inaccurate due to the lag period between growth in basic activities and the resulting service growth. Sixth, the multiplier is an average for all activity, it does not apply to or reveal any detailed information on any one industry. Seventh, economic base theory considers only export activities as basic. It assumes that growth in basic activities will result in a growth in service activities. However, the service capacity to grow may be filled, i.e., the market full, and therefore no significant service growth is going to occur regardless of basic growth. Finally, export base theory makes no distinction between long run economic growth and short run fluctuations. Multiplier analysis usually applies only to short run fluctuations. The economic base theory is a concept of only short run income determination. An accurate analysis of economic growth (long-run phenomenon) requires consideration of many diverse factors, some of which were mentioned in section II. 5. B.

II. 8. Conclusion

Despite the numerous disadvantages and limitations of an economic base analysis, such a study can provide a great deal of information regarding the structure, strengths and weaknesses of the local economy. In addition, the method is a relatively inexpensive tool of regional analysis. When applied with a full understanding of its problems and limitations, the economic base study can be a vital aid to planners, government officials and business leaders.

The following sections present several modifications of an alternative approach to economic analysis and projection, Input-Output Analysis. The main advantage of this procedure over the economic base approach is that it permits various detailed components of a regional economy to be analyzed and their growth estimated. While the economic base approach ignores the interrelationships among industries, and thus provides no information of the effects of an increase or decrease in output of a given industry upon the rest of the economy within an area, the input-output approach is quite adept at identifying inter-industry links and analyzing their impacts upon regional growth.

III. Input-Output Analysis

III. 1. General Description and Theory

As stated in the preceeding section, the economic base multiplier approach assumes that local output and employment will grow only if export activity grows. Input-Output (I-O) analysis is also used to estimate such economic impacts, but its distinctive nature derives mainly from its detailed analysis of individual industries within the local economy. This approach identifies the complex interindustry relations between purchases of inputs and the distribution of outputs. I-O analysis identifies and tabulates the flows of products from each producing sector to each purchasing sector. Generally such flows are measured for a given time period (usually one year) in terms of dollar sales. By means of such detailed tracing of output flows, the expansion and contraction effects of each industry upon the local economy can by analyzed.

In addition to a detailed identification of the interindustry input structure and output distribution for each industry, I-O analysis examines components of the "final demand" sector. By disaggregating exogenous final demand into components such as export, government, local investment, inventory change, households, etc., the analyst can better estimate the impact upon the local economy resulting from an increase or decrease in each component of final demand. Thus one would expect to obtain more accurate estimates than is provided by a gross aggregation of final demand into one component as done in an economic base analysis.

Figure 1 is a hypothetical I-O transaction table. In this example there are 5 endogenous sectors and three exogenous final demand sectors. The numerical values represent gross dollar flows among sectors, for example, agriculture purchased $10.00 of input from the transportation sector to produce its total output of $100.00.

TO / FROM		Agri. 1	Mining 2	Eng. 3	Bus. Serv. 4	Trans. 5	Households 6	Govt. 7	Exports 8	Total Output 9
Agriculture	1	21	-	9	3	-	30	5	22	100
Mining	2	1	8	7	29	-	25	2	23	100
Engineering	3	3	20	-	50	7	5	4	6	100
Business Services	4	31	2	38	-	3	12	-	1	100
Transportation	5	10	25	26	1	4	9	-	5	100
Value Added	6	20	40	10	17	40	2	4	-	
Imports	7	14	5	10	-	46	77	-	-	
Total Output	8	100	100	100	100	100	160	15	67	

Figure 1 Gross Flows

The Input-Output model, therefore, is an extremely useful descriptive device because it (1) concisely records, in an internally consistent manner, a large amount of information on the sectoral interrelations within a local economy; (2) reveals gaps in existing data, such as the increased income to households due to additional output in a given sector; and, (3) permits an analysis of the impacts upon the local economy of both existing and new, proposed industries.

III. 2. Uses

The potential of I-O analysis as a descriptive device has been

briefly mentioned in the above section. However, the greatest advantage of this approach lies in its estimation of various impacts upon the local economy resulting from fluctuations in demand for the goods and services produced in the area's economy. Several standard uses, as well as some which are relatively new, are listed below. This list is by no means inclusive, and is continually being expanded by on-going innovative research.

(1) Direct input requirements - The A coefficients matrix, a_{ij}, describes the direct input requirements per dollar of industry j output, from industry i, thus identifying the linkage between the output of a given industry with the output of all other industries within the economy. These are found by dividing the input flow from i to j by j's total output. Figure 2 shows the a_{ij} coefficients derived from the hypothetical model shown in Figure 1.

sectors	1	2	3	4	5
1	.21	-	.09	.03	-
2	.01	.08	.07	.29	-
3	.03	.20	-	.50	.07
4	.31	.02	.38	-	.03
5	.10	.25	.26	.01	.04

Figure 2 A - Matrix (a_{ij})

(2) Direct and indirect input requirements - the coefficients in the Leontief inverse matrix $(I-A)^{-1}$ describe both the direct requirements and the additional requirements necessary to sustain production of the required inputs. The Leontief inverse matrix of

the present example is shown in Figure 3.

(3) Traditional multipliers

(a) Output multipliers - Estimates of the total direct, and direct and indirect production required of the entire economy to meet the requirements nesessary for a $1 increase in sales by industry j are obtained by summing the j column coefficients in the (A) and $(I-A)^{-1}$ matrices respectively. The direct plus indirect output multipliers are also given in Figure 3.

		sectors				
		1	2	3	4	5
	1	1.33	.05	.18	.15	.02
	2	.23	1.17	.30	.50	.04
sectors	3	.40	.36	1.41	.82	.13
	4	.58	.19	.61	1.38	.09
	5	.31	.41	.48	.38	1.09
output multipliers		2.85	2.18	2.98	3.23	1.37

Figure 3 $(I-A)^{-1}$ Matrix

(b) Income multipliers - By making the households sector endogenous, purchases from households as labor inputs, and inputs to households as consumption, can be considered. A family of income multipliers describing direct, and direct, indirect and induced income to households per dollar of final demand are obtained. In addition, multipliers can be derived which describe per dollar of direct income, the direct and indirect income generated (Type I), and the direct, indirect and induced income generated (Type II). The Type I multiplier ignores induced spending of income, i.e.

assumes the marginal propensity to consume equals zero, and thus underestimates the income effect. The Type II multiplier includes induced spending, assuming the MPC equals the average propensity to consume, and therefore may overestimate income generated. Several approaches which attempt to solve these estimation problems are discussed in detail within much of the I-O literature, and should be referred to when attempting to interpret and apply the income multipliers.

In figure 4 the households endogenous coefficients are shown for the A and $(I-A)^{-1}$ matrices.

		sectors					HH
		1	2	3	4	5	6
sectors	1	.21	-	.09	.03	-	.18
	2	.01	.08	.07	.29	-	.15
	3	.03	.20	-	.50	.07	.03
	4	.31	.02	.38	-	.03	.07
	5	.10	.25	.26	.01	.04	.05
HH	6	.20	.40	.10	.17	.40	.01

(4a) A* Matrix

		sectors					HH
		1	2	3	4	5	6
sectors	1	1.59	.34	.43	.43	.21	.41
	2	.48	1.46	.54	.78	.23	.41
	3	.62	.61	1.62	1.07	.30	.36
	4	.82	.47	.84	1.65	.28	.39
	5	.52	.65	.68	.62	1.25	.34
HH	6	.93	1.07	.89	1.05	.72	1.50

(4b) $(I-A^*)^{-1}$ Matrix

Figure 4 Households Endogenous

Calculation of the various income multipliers, using sector 1 as an example, is shown below.

(1) Direct income change (household now coefficient - a_{ij})	0.20
(2) Direct plus indirect income change per dollar of final demand	0.62
(3) Type I multiplier (2) ÷ (1)	3.10
(4) Direct, indirect and induced income change (household coefficient in $(I-A^*)^{-1}$ matrix)	0.93
(5) Type II multiplier (4) ÷ (1)	4.65

(4) Projection - Impact analysis - Certainly the most useful aspect of I-O analysis is its ability to examine the impact of projected final demand changes upon various components of the local economy.

(a) Output projection - By premultiplying a column vector of projected final demand for year (t) by the (A) and $(I-A)^{-1}$ matrices, estimates of direct, and direct and indirect output required to meet this demand are obtained, respectively. Thus the analyst is able to determine future levels of output for each sector within the economy.

Using the above example, assume that the aggregate final demand for the following year has been estimated to increase as follows.

Agriculture	+$200.00
Mining	0
Engineering	0
Business Service	0
Transportation	+$100.00

The total change in output (direct plus indirect) that is required

from each sector to meet this estimate change in final demand is:

	sector 1	2	3	4	5					
1	1.33	.05	.18	.15	.02		200		268	Agri.
2	.23	1.17	.30	.50	.04		0		50	Mining
3	.40	.36	1.41	.82	.13	X	0	=	93	Eng.
4	.58	.19	.61	1.38	.09		0		125	Bus. Serv.
5	.31	.41	.48	.38	1.09		100		171	Trans.
			$(I-A)^{-1}$				estimated change in final demand		total change in output	

Thus the indirect ties become apparent in that even though there is no estimated increase in the mining sector final demand, that sector must increase output by $50.00 to meet the agriculture and transportation final demand increases. This is due to the fact that the mining sector (as do the others) supplies inputs necessary for sectors (1) and (5) to increase their total output.

(b) Income projection - By inclusion of the households sector in the (A) and $(I-A)^{-1}$ matrices the induced output, i.e., that generated by respending of income, required to meet final demand can also be estimated. Also through use of the above mentioned multipliers, estimates of the total income to the area resulting from changes in final demand can be generated for each industry within the local economy.

(c) Employment projection - A matrix of employment coefficients, i.e. total employment/output, describing the required inputs of

employment for each sector can be constructed and post multiplied by projections of total output to obtain estimates of the total employment generated by such output. Using the above estimates of rural output, the employment increase in each sector which results from increased output is found by;

Agri.	.21	0	0	0	0		268		42
Mining	0	.05	0	0	0		50		2.5
Eng.	0	0	.50	0	0	X	93	=	46.5
B. S.	0	0	0	.36	0		125		45
Trans.	0	0	0	0	.17		171		29
	direct plus indirect employment in j required per dollar of j's output						total output		total employment increase by sector

To these employment projections a ratio relating employment to population can be applied to obtain estimates of future population within the area. These projections can then be compared with population estimated by conventional means such as mathematical extrapolation or the age-cohort survival technique. Employment projections may be further disaggregated into occupational groups (see Miernyk 1970). This involves the construction of a matrix of total employment in each industry broken down by occupational type. This new matrix may be used to estimate the types of jobs resulting from changes in final demand. Such a breakdown has several implications for determining the types of educational and training programs needed in an area to meet future manpower requirements. For example, the Miernyk report (Miernyk 1970) projected a significant portion of

the new employment created in the West Virginia economy to be white collar and technical jobs. Although some of these jobs will be filled by immigrants, this estimate does provide some guidance in determining the types of training programs which the state's educational system should be providing to meet future employment requirements.

(d) Resource requirements projection - Estimates of the ecological inputs (land, water, etc.) required to meet projected output can be generated through an I-O analysis. This is done by constructing a matrix of inputs/total output coefficients for each industry and post multiplying this matrix by projected total output. Although many of the ecological inputs-economic output relationships at present are extremely approximative, such a projection can provide the planner and decision-maker with general estimates of the demand placed upon the ecological system by changes in final demand. Some examples of models of this type are discussed in the following sections.

(e) Environmental pollution projection - The Isard (Isard 1972), Hite and Laurent (Hite 1972), and Korbach and Cumberland (Korbach 1972) reports have attempted to extend the resource projections to estimating the various pollutants resulting from increased output within the local economy. Coefficients describing the amount of each pollutant (e.g. BOD, SO_2, particulates, etc.) per dollar of output for each industry are placed in matrix form and applied to projected output to obtain such estimates. This type of an analysis

has, obviously, tremendous potential for an area concerned with preserving environmental quality. Comparison of the estimates with Federal pollution emission standards can aid in identifying both the industries which will be required to provide treatment or control technology and those which may have to be prohibited from expanding or even locating within the area. The Hite and Laurent study contains a statistical model, (although certain problems with its use do exist) for estimating the area residents' "willingness-to-pay" for various levels of environmental quality e.g., an air quality level with no perceivable odor or soiling. This dollar figure willing to be paid to achieve such a level may be compared with the local income and pollution generated by the growth or location of a particular industry, to determine the total benefits and cost (i.e. total income - treatment costs) resulting from the economic activity. Although a projection of generated pollutants is extremely valuable, present knowledge and data permit only approximate estimates. However by identifying existing data gaps, research can be initiated which will eventually provide accurate estimates of the relationship between pollutant emissions and economic output.

(f) Interregional analysis - I-O analysis can and has been applied to studying structural relationships among a number of regions. However, the errors resulting from the various limiting assumptions as discussed in Section III. 3., are compounded when the study is extended to an interregional analysis. For example, price and trade pattern changes are certain to be greater over a larger study area.

Whether or not an interregional approach is necessary, depends upon the economic characteristics of the area and the objective of the study. Errors resulting from not considering output flows to other regions are relatively small when the study area itself is large, diversified and relatively self dependent. For those regions which are strongly linked to other regions, i.e., they receive a substantial number of their inputs from outside of the study region, ignoring the interregional feedbacks may result in serious estimation errors. Similar errors may also result when the study region supplies a large portion of inputs to industries and consumers in other regions. When there is no consideration of the effects of output and consumption, expansion and contraction in other regions upon the study region, the regional multipliers may be meaningless. Thus a knowledge of the area can provide the analyst some guidance in determining the importance of interregional consideration.

(g) Policy evaluation - The above mentioned projections available through an I-O analysis suggest that such a study can be extremely helpful to the decision-maker concerned with public policy evaluation. Once a clear set of regional goals has been identified, the various impacts of future economic development can be analyzed and the alternatives identified. For example, if a public goal is to provide maximum employment, those industries identified as providing the greatest number of jobs resulting from growth or location in the area may be determined most desirable and therefore should be

encouraged. Similar approaches to goals of increasing local income, preservation of environmental quality, or efficient resource utilization may be applied. Proposed investment in human resources, schools, hospitals, social services, etc., may be examined to determine whether they will adequately meet projected output, population increases. Obviously numerous other policy implications exist, depending largely upon the purpose of the study, the degree of detail employed within the model, and the goals and characteristics of the study area.

III. 3. Limitations of Using Input-Output Analysis for Projection

(A) Data Requirements - The data required by the I-O transaction table are enormous, as would be expected by the information which it provides. For greatest accuracy such data should be primary, i.e. collected through personal interviews and questionnaires. This can only be accomplished with considerable expenditure of time and money. Adjustment of national data, which are available from national I-O studies, is one approach to overcoming the cost problem, the accuracy of such an approach obviously depends upon the regional similarity to national averages (which the national coefficients are), and the type of adjustment made. While it may be possible to obtain accurate estimates of the regional technical coefficients, estimation of regional purchases and imports remain critical, unresolved issues. Several other approaches to overcoming the cost problem include: (i) consideration of only "basic" industries (Bonner 1967); (ii) use of employment rather than output as the unit of measure (Hanson &

Tiebout); and, (iii) extreme aggregation of sectors (Doeksen & Little 1968).

(B) Assumption of constant coefficients - When using the I-O matrices for projection purposes, one makes the assumption that the various coefficients remain constant over the projection period (there have been exceptions: some of these are mentioned below). There are several factors which seriously question the validity of such an assumption, and are discussed below.

(B.1) Technological change - As the projection period increases, so does the likelihood that research and development will produce changes in existing production technology, and thus possible changes in the type and amount of inputs required. The Miernyk study (Miernyk, 1970) attempts to account for technological changes by assuming that existing "best practice" (most efficient) technology will, in time, become "average technology." Thus the input coefficients of best practice industries are linearly interpolated over a period of years such that the A matrix at time t + m (the projection period) contains the coefficients of the best practice firms observed at time t. In a sense then, the best practice technology becomes the average technology over a period of several years.

(B.2) Input substitution - It is possible that firms will find new inputs to replace those presently being used, the degree depending greatly on the type of output produced.

(B.3) Economies of scale - As production of a good increases it is not necessarily correct to assume that inputs will increase proportionally. As the size of production increases, certain economies may result which require considerably less inputs/unit of output.

(B.4) Agglomeration economies - Agglomeration of industries which utilize the same or similar inputs may result in a decrease in the relative price of these inputs. In addition such agglomeration can affect output flows among these industries.

(B.5) Relative price changes - Market conditions can drastically change input prices (e.g. oil) in a relatively short period of time. In addition to changing the coefficient values, such a price change may result in the amount of inputs purchased and output consumed by final demand.

(B.6) Product mix changes - Depending upon the degree of industrial disaggregation used, the input coefficients describe the input requirements necessary for the production of an "average" good. For an industry producing many products, changes in the amount of each type of good produced can drastically alter the industry's average input requirements.

(B.7) Trade pattern changes - The sources, and possibly prices, of inputs may change over the projection period, thus changing the structural flows described by

the A matrix coefficients. Miernyk attempts to account for such effects by examining past trade pattern changes as a basis for projection of future changes.

(B.8) Final demand estimates - A major problem is the estimation of all relevant final demand sectors. As mentioned earlier, these usually include households, government spending, local investment, foreign exports, inventory change, and capital formation. One must also attempt to accurately estimate changes in these scctors. Typically, estimates are made by: (i) an assumed constant share of existing national projections; (ii) comparison with demand experienced by a region whose past structure is similar to the study region's present structure; (iii) examination of past changes within the study region; and, (iv) estimation based upon knowledge of local business leaders. Obviously projections of output, employment, income, etc. will reflect errors in the final demand estimates upon which they are based.

III. 4. Summary

Despite the "static" nature and the limitations mentioned, I-O analysis has been demonstrated many times to be an extremely useful technique for examining the economic structure of an area. Any projection, regardless of the technique involved, will be imperfect in a dynamic society. However, when a projection is made within the detailed, systematic and consistent framework of an I-O table, the

analyst is less likely to overlook important considerations which may affect the accuracy of a projection. It therefore seems entirely reasonable to use the I-O technique for projection purposes as an approximative procedure. Because of the nature of the I-O coefficients, it follows that the accuracy of the projections will tend to be greater the closer the projection year is to the base year. Consequently the more frequently an I-O table is revised and updated, the more likely will the projections approximate actual growth conditions.

As mentioned earlier, numerous public agencies and business groups can benefit from the information provided by I-O analysis. A cooperative effort among these groups would minimize the costs in time and money which each must individually bear in order to garnish such benefits. Therefore, the enormous requirement of time and money would not be an insurmountable problem, forclosing consideration of conducting a detailed I-O analysis for the State of Illinois.

In the following section three hybrids of the standard I-O technique are reviewed. These smaller, aggregated models have been developed in response to overcoming the extensive time and money expenditures required by the larger, detailed models. For certain types of studies and regions, the hybrid models may be capable of providing reasonably acceptable economic projections at considerably less costs.

IV. Hybrid Approaches to Input-Output Analysis

IV. 1. Introduction

Several hybrids of the conventional I-O technique have been proposed in an effort to overcome the enormous amount of data required by the conventional study. The time and money expenditure necessary for such studies has prohibited employment of the I-O technique for smaller urban areas which do not possess unlimited research resources. This has resulted in much of public decision making for such areas being based either upon the most elementary analysis and projection techniques, or entirely upon subjective opinion.

Public and private decisions, however, will be made, whether or not accurate estimates of their economic impacts precedes these decisions. The three techniques discussed below are examples of the on-going effort to make the necessary detailed information available to local planners and decision-makers for examining and estimating the probable impacts of alternative courses of action. These simpler techniques are modifications of the standard I-O study in terms of: (1) the amount of data collected; (2) the types of coefficients used; and, (3) the size of the transaction table. The following discussion is aimed at describing the incorporation of the techniques and the reliability of the estimates they provide as compared to conventional I-O projections. For a complete description of the procedure involved in conducting a particular study, the reader is referred to the original published report.

IV. 2. Technique for Area Planning (Bonner, 1967)

Because of enormous cost and time constraints, a detailed interindustry analysis, using I-O methods, has seldom been carried out for small urban areas. However, the benefits gained through an analysis of alternative development policies are just as valuable and important for the small urban areas as they are for large metropolitan regions. Every public decision by a policy-making body will have some impact on the local enomomy: the immediate impacts are often readily visible. However, secondary impacts are often unknown due to lack of local economic analysis. In an effort to overcome the costly construction of a detailed I-O table for economic analysis, the Technique for Area Planning (TAP) method has been proposed. This method is a less costly, simplified I-O technique which has great promise for policy analysis of small urban regions. This technique eliminates much of the transaction table data ordinarily collected through surveys. Detailed transaction data are not collected from firms in sectors where reasonable estimates can be made using more limited information.

IV. 2. A. Overview of TAP; major and minor sectors

The TAP procedure identifies two classifications of sectors: major and minor. Major employers, especially in small areas, are those upon which the prime economic growth of the region rests. For small areas these are usually a few large, easily identifiable firms. The information collected for major sectors is essentially identical to that collected for a conventional I-O study, but

because of their small number, the detailed information can be collected without enormous expenditures.

Minor sectors include those which are assumed to contribute relatively little to the area's growth. The data for these sectors are reduced to a few simple items which can be easily obtained, thus eliminating much of the costly data collection of the traditional I-O study.

The TAP procedure is basically a compromise between the traditional I-O study and the economic base technique. Data for the minor or non-basic sectors are highly aggregated. As in an economic base study the definition of final demand items, other than exports receives only secondary consideration.

IV. 2. B. Suitability of an Area for TAP Analysis

Tests of the TAP approach have shown that as regions become larger and more diversified, approximations obtained with TAP become less accurate. These tests have indicated that the best prediction of a region's suitability for TAP usage is the Index of Specialization. This index approximates the degree of economic specialization or diversity of the economic sectors within the region. A minimum value of zero indicates complete diversity while a maximum of one indicates high specialization of firms in one or more economic areas of production. This index is calculated by subtracting each sector's percentage distribution of employment in the region from that sector's corresponding employment percentage

in the nation, summing either the positive or negative numbers, and dividing by 100. For example, consider a region with 3 sectors. . .

	% employment in region	% employment in nation	difference
Sector 1	46	40	-6
Sector 2	24	38	+14
Sector 3	30	22	-8

Index of Specialization = .14, indicating a rather diverse economic structure. The available case studies suggest that regions with an index lower than $\pm$0.20 are not suitable for TAP analysis. This would follow a priori expectations that the larger and more diverse a region is (thus a low index) the greater the number of sectors that would apt to be given only secondary consideration in the TAP procedure. The greater the number of potentially important sectors which TAP "ignores", the greater the likely errors obtained in using the procedure. Comparison of TAP application to regions of varying size will be discussed later in the report.

IV. 2. C. Identification of Major Firms

The analyst begins by classifying firms into major and minor categories by use of a simple economic base approach. This type of data collection can aid in tentatively identifying those sectors which represent the main growth stimulus of the region.

A major firm is obviously extremely important to the study area in a number of ways. First a major firm is one which employs a large percentage (greater than 10%) of the local labor force. In

addition, a major firm will be likely to sell most of its product outside the study region, thus causing money flows into the area. Use of location quotients can aid in identification of sectors which sell outside of the study region. A location quotient greater than 1.0 generally indicates a sector whose production exceeds that needed for local demand and thus in part is exported. Location quotient analysis is based upon a number of limiting assumptions and the reader should refer to the various written critiques of the method before attempting to employ it for purposes of identifying major sectors; for instance, see: (Tiebout, 1962).

Another consideration which may determine a major firm is whether or not the firm is growing or declining relative to other firms in the region or nation. Clearly, such growth patterns are important to the region's economy. Finally firms should be studied in terms of the demands which they place upon local raw materials, supplies, labor, etc., although, a priori, this may be difficult to estimate. TAP analysis defined minor firms as those not listed as major.

IV. 2. D. Data Collection

The tremendous cost advantage of the TAP procedure over a conventional I-O study is due to the amount of and means by which data are collected. Those savings must be weighed against the loss of detailed information on many sectors and the increased probability of error as the size of the study region increases.

Through personal interviews, complete data on sales and purchases for each major firm are obtained. These include sales and purchases with other major firms, sales to minor sectors, households, and final demand, and purchases from minor sectors, imports and total payments. This survey of major firms should follow standard interviewing procedures of letters of introduction, interviewer familiarity with the production processes, and the assurance of confidentiality.

Minor sector data consisting of sales to exports, purchases from imports, and households, and total output, are obtained through a brief mail questionnaire. Firms receiving the questionnaire are selected through a stratified random sampling procedure. As always, care should be taken so that the selected sample is representative of the entire sector from which it was drawn.

Because considerably less data are collected, a number of the cells in the transaction table remain blank. Through the household row and column, the table is adjusted for this missing data. Minor sectors' sales are allocated to major sectors, households, and final demand. Sales to households are determined by subtracting the sum of sales to major sectors and final demand from total minor sector gross output. A conventional I-O study would require data collection which would permit distribution of this amount to interindustry transactions among minor sectors - an additional reason why smaller, highly specialized regions (i.e. few minor firms) are better suited to the TAP procedure.

A somewhat similar procedure is used when a household row

intersects each minor sector column. Known minor sector purchases from major sectors, households, imports and other payments are summed and subtracted from total gross output. The resulting figure represents minor sector purchases from other minor sectors. The amount for each minor sector column is adjusted by the fraction $(1 - \frac{\text{minor sector imports}}{\text{minor sector total outlays}})$, added to the amount of known household purchases and recorded in the appropriate household row cell.

Further adjustments of the data using control totals, margins, etc. are also applied to achieve a final balanced, modified transaction table. As with the above described adjustment, these procedures are applied to reconcile differences and distribute the major and minor firm transactions. For a complete description of the necessary adjustments the reader should consult the actual TAP report.

IV. 2. E. Projection of Economic Conditions

As with the conventional I-O table, the modified transaction table can be inverted to obtain direct, indirect and induced technical coefficients. Knowledge of regional demand and projections of national economic activity are then used to estimate changes in sales to final demand. These changes, based upon the relationships described in the modified table, are then used to estimate changes in production for each local sector.

Finally employment coefficients, adjusted for expected changes in labor productivity may then be computed and used to estimate total

changes in employment due to increased output in the region. From these expected increases in employment, population projections can be made, thus giving some estimate of the demand for housing, utilities and other community facilities.

These estimates may be used by local decision-makers in evaluating the impacts of proposed development policies. By examining the total impact, the decision-maker can better identify which existing firms should be encouraged to expand, or which of several new industries should be encouraged to locate in the area. The technique enables one to look beyond the immediately observable effects of economic development. It enables identifiaction of not only the new jobs created in a new firm, but also the new jobs in existing firms that result due to increased production needed to supply inputs to the new industry. Obviously proper evaluation of alternatives necessitates an understanding of community objectives. A clear policy statement of what is to be achieved will enable a better selection of the most desirable alternative. If, for example, total increase in employment is the only community objective the alternative which results in the largest total employment increase is to be preferred.

IV. 2. F. Limitations of the TAP Procedure

Because the TAP method is a compromise between the conventional I-O study and economic base analysis, it is subject to the limitations which both techniques possess. In addition, the information provided by the procedure is limited due to the extreme aggregation of the

sectors. Because many of the minor sector transactions are simply ignored or allocated to households, final demand, etc., additional potential sources of error are present within the procedure. The conventional limitations of the I-O technique (such as the assumption of constant trade patterns) may be compounded due to the aggregation or exclusion of certain sectors.

IV. 2. G. Comparison of TAP Multipliers and Conventional I-O Multipliers

As stated earlier, the suitability of an area to TAP analysis has been found to be closely related to the Index of Specialization. A comparison of TAP sector multipliers with those obtained for a conventional I-O table have been made for seven regions of varying size (14 - 34 sectors). For relatively small specialized regions (i.e. Index greater than .25) the impact estimates obtained with the TAP procedure were generally within five percent of estimates made with the traditional I-O method. Since even the most complete I-O study provides estimates which are subject to error, most small communities are expected to find this degree of accuracy well within acceptable limits.

For the smaller regions the TAP estimates are slightly larger than the conventional estimates. However, as the region increases in size and diversity, the TAP multipliers are as large as 1½ times the standard I-O multipliers, thus indicating the need for a more detailed analysis and disaggregation of sectors within the economy. For the smaller regions where available resources are limited, the

TAP procedure does offer a helpful means of estimating economic impacts. Additional investigation and improvement of the method will continually increase its usefulness to the decision-making process for small urban areas.

IV. 3. Effect of Size of the I-O Model on the Results of an Impact Analysis (Doeksen and Little 1968)

Another approach to overcoming the problem of costly data acquisition is based on the contention that a small aggregated model will yield specific sector multipliers similar or nearly identical to those obtained from a large disaggregated model. The hypothesis states that the multiplier for a specific sector will be approximately the same regardless of how many total sectors are in the model.

IV. 3. A. Testing of the Hypothesis

Multiplier comparisons were made using four empirical models by Doeksen and Little (1968). Two hypothesized 25 sector models, and two actual models, one a 27 sector model of Washington state and the other a 29 sector model of Oregon, were analyzed. Three sectors in each model were randomly chosen to remain unaggregated. The remaining sectors were then aggregated two at a time, until the model was aggregated up to a total of four sectors - 3 sectors unaggregated and one sector representing the remaining sectors.

IV. 3. B. Empirical Results

A comparison of the multipliers for the three sectors before and after aggregation revealed several conclusions. First, the

amount of change in the multipliers was largely determined in the initial size of the multiplier before aggregation. Smaller multipliers in general increased in the third decimal place as the model size decreased. Second, the larger multipliers tended to decrease as model size decreased. Third, a slight increase in income multipliers resulted as model size decreased. The results of the comparison revealed that all changes were small, supporting the hypothesis that model size has little effect on multiplier size.

To further support the hypothesis, the size of the model was regressed against the size of the multiplier. The resulting regression coefficients were all approximately zero, indicating little or no relationship between the two variables. Finally, the coefficient of variation of each sector's multiplier as the model was aggregated, was calculated. These coefficients indicated a small variation in each multiplier, again supporting the original hypothesis.

Additional study on the aggregation of sectors (Hewings 1972) has shown that the resulting multipliers are suitable for impact analysis only for those sectors which remain unaggregated. The aggregated sectors' multipliers do significantly change, and, therefore, should not be used to estimate impacts. This simply suggests that for studies which are concerned with examining the influences of a few specific sectors, those sectors must remain unaggregated and only their multipliers used for analysis.

IV. 3. C. Implications and Conclusions

The hypothesis that there is no relationship between multiplier

size and model size is strongly supported by the empirical results of the Doeksen and Little (1968) article. From this, several implications may be posed. First, if the available data and resources are limited, construction of a small model will yield impact estimates similar to a large model. Thus the impact of an existing or new industry, with a high degree of aggregation of the remaining sectors of the economy. From this small model relatively accurate estimates of the impact on output, employment, and income may be obtained. By adding a row and column for each alternative industry being considered, the relative impacts of each may be compared.

The above technique is very similar to the TAP procedure in that it contends that highly useful estimates can be obtained by constructing a model with separate sectors for each major industry, and aggregating the remaining sectors in broad classifications. Certainly the implications of such an approach are promising and warrant further research to substantiate the conclusions drawn in the study.

IV. 4. An Intersectoral Flows Analysis of the California Economy (Hansen and Tiebout, 1963)

The basic objective of the study was to develop a model which is operational in that sense that the necessary data could be obtained at a reasonable cost. This model also incorporates certain features of the standard I-O analysis and base multiplier approach.

IV. 4. A. The Model

The empirical model examines only the flow of outputs,

disregarding input requirements. Using employment rather than sales as the unit of measurement, the model divides the total output of industry i into flows to other j industries in the region and flows to final demand sectors. This results in an interindustry table in that the output (employment) of one industry becomes the input to another industry, though only local inputs and local industries are considered. The model recognizes, however, that all employment assigned to other industries becomes indirectly assigned to the final demand sectors, i.e., employment in all industries is tied directly or indirectly to final demand. As stated earlier, the induced effects are estimated by making certain final demand sectors endogenous. Although this may be done for all sectors, it is typically carried out with the consumption or households sector.

IV. 4. B. Data Sources

Even with this modified I-O approach, data requirements are somewhat severe. Data tracing sales flows from one industry to other industries and the final demand sectors are necessary. Data collection is based upon the assumption that firms better know the destination of their outputs, than do they know the origin of their inputs. The reasoning being that inputs are so diverse and varied that information of their origin is much more difficult to obtain.

Manufacturing data were obtained through the use of a relatively simple mail questionnaire for a sample of firms. The questionnaire asked only information on the breakdown of a typical dollar of sales

to local industries and final demand. The results were then weighted according to firm size, and aggregated on a 2-digit basis.

Because of the large amount of data already available, the questionnaire approach was not employed in acquiring data for the non-manufacturing sectors. These data included, for example, commodity reports for agriculture, and building permits and awarded contracts for the construction sector.

The final accuracy and suitability of the method rests on the accuracy of the initial data. No comparison is made in the article between employment multipliers generated with this approach and those derived from a conventional I-O study. Obviously, aggregation at the 2-digit level will reveal no information on individual industries contained within a sector. In addition, collection of data solely on an employment basis forecloses the use of various published control totals for the balancing and identification of certain inconsistencies in the data. For example, the absence of information on inputs excludes the possibility of an independant check on output flows. Also, there are the usual accuracy problems inherent in extending sample data to represent an entire sector.

IV. 4. C. Determination of Employment Multipliers

IV. 4. C. a. The Initial Allocation

The employment data for industry i is initially allocated to each local industry and the final demand sectors in accordance with

i's sales data obtained through the questionnaire or other published sources. For any given industry, summing the employment allocated to all final demand sectors will give a measure of the amount of that industry's employment directly tied to final demand sales. Also, from this initial allocation employment input coefficients can be derived. These are determined by dividing the number of jobs in i related to industry j, by the total employment in j. The resulting coefficients represent the amount of employment in i required per employee in j - similar to the standard I-O coefficient, except for the unit of measure.

IV. 4. C. b. The Final Allocations

Because the objective of the study is to trace employment to final demand sectors, additional allocations are necessary. This is accomplished by reallocating to final demand the inputs from i to j in the same manner as j's employment is allocated to the final demand sectors. Thus if 4% of j's total employment goes to the government final demand sector, then 4% of the employment input from i to j would also be allocated to that sector. Employment from i to j is allocated in a similar manner to other industries, based upon j's employment allocation to those industries. The latter allocation thus results in some of i's employment still being related to other local industries. However, by continually carrying out this iterative process, all employment in i originally allocated to local industries can be indirectly related to the final demand sectors.

The resulting table gives an estimate of each sector's direct and indirect employment which is linked to each of the final demand sectors. From this distribution of employment to final demand, the impacts on employment due to changes in the various final demand sectors can be estimated. Summing the employment allocated to each final demand sector for each industry gives an estimate of the total employment in the entire region which is tied to each export sector. These employment multipliers can then be applied to estimate the effects of changes in final demand upon the regional economy. In addition, the impact of the employment of i due to increases in j's output can be estimated earlier in the allocation procedure, that is, before employment is totally distributed to the final demand sectors.

IV. 4. D. Conclusions

As stated earlier, most of the limitations of this approach center around the type of, and means by which the data are collected. However, the main advantage of this approach is its operational simplicity. The authors contend that reasonable estimates of the employment multipliers may be obtained with this method at a realistic cost. Much of the reliability does rest upon the relative size of each sector being considered. When the employment of very small firms is allocated to other industries and ultimately to final demand, the coefficients can be so extremely small as to be meaningless.

The hybrid models discussed in this section would seem to possess

several clear advantages over economic base models. The smaller demands on data requirements further serve to strengthen these advantages especially in comparison to full-scale I-O models.

V. Economic-Ecologic Linkages Models

V. 1. The Need for such Models

The preceeding chapter discussed alternative approaches to analyzing and estimating the economic output, population, and income growth of a particular area. The valuable information gained through such techniques as economic base and I-O analysis has been suggested in this report and in numerous other publications. Through an identification of an area's interindustrial structure, the impact of changes in output (final demand) upon the growth of an area can be estimated. Such estimates can aid planners, business leaders and public decision-makers in evaluating whether alternative policy proposals will result in achievement of various public goals.

The regional analysis techniques noted earlier have been used almost exclusively for the purpose of estimating the economic impacts of expanded output. However, the assumption that the benefits of development will exceed the cost is, in many instances, no longer tenable when the ecologic impacts are included in the analysis. No longer can regional development and regional planning be considered in their traditionally narrow contexts, with emphasis on the strict economics of the development and only a cursory consideration of the natural environment. There is a growing realization that effective regional planning and resource management requires a synthesis of the analysis of two systems, the economic and ecologic systems. Regional development proposals must now be viewed in such broader contexts as air and water

pollution, solid waste disposal, and open space preservation. The problem that economic growth often brings with it heavy environmental costs must be confronted.

The once universally accepted regional goal of increasing exports must now be weighed against the realization that the various economic benefits are obtained at the cost of decreasing environmental quality. By producing goods and services for export, a region generates an internal flow of wastes in return for an inflow of credits. Conversely, when a region imports for delivery to either the intermediate or final demand sectors, it generates waste flows in other regions in return for an outflow of credits. Thus, at the possible cost of jobs and incomes, environmental quality in the region is maintained. However, wastes are generated at the consumption as well as the production stage. Because all goods ultimately end up as wastes through the process of consumption, a region cannot completely avoid the large waste problem caused by goods production by simply importing those goods. Through processing and packaging, the production sectors generate their wastes immediately. Wastes generated through the comsumption processes are usually associated with a time lag. Disposal of comsumer non-durables occurs relatively rapidly, while durables may require several years to become wastes. Therefore effective analysis of the environmental ecological linkages requires consideration of the time problem as well as the economic production problem.

V. 2. Ecologic-Economic Analysis for Regional Development (Isard 1972)

Certainly one of the most ambitious efforts to synthesize analysis of the economic and ecologic systems was the model first presented by Walter Isard at the Regional Science Association meetings in 1967, and later applied to a specitic site, the results being published in 1972. Figure 1 illustrates the general format of the Isard model:

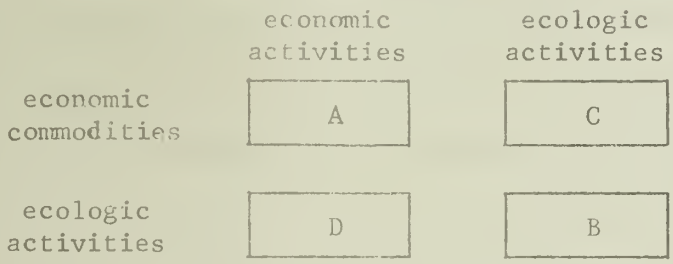


Figure 1

where: A = standard interindustry I-O matrix

B = similar inter-ecological system matrix

C = exports from economic system to the environment

D = inputs to economic system from the environment.

Thus, the Isard model attempts to describe ecologic inputs to the economic system, economic outputs (pollutants) to the environment, and inter-ecological system relationships. Isard's model contains an extremely large number of disaggregated ecological sectors. The all-encompassing nature of the model makes such disaggregation necessary and, as a result, requires enormous amounts of data, which, given the present status of ecological science, is largely unavailable.

An example will serve to demonstrate the model's all-encompassing nature. The Isard classification scheme divides the ecologic sectors into two main groups: abiotic and biotic. The abiotic grouping contains: climate, geology, physiography, hydrology, and soils. The biotic grouping is organized into plants and animals. One can readily see that if the researcher attempts to identify only the linkages of these categories to the economic system, he will soon be overwhelmed with the data requirements. This is further complicated by attempting to analyze the inter-ecological relationships as Isard suggests in his report. The report, for example, traces the winter flounder's food chain, identifying the interrelated components such as algae, detritus, mollusca, and crustaceae. Similar to the interindustry format, coefficients describing the numbers of crustaceae/pound of winter flounder are sought for the analysis. As this type of consideration is extended to every significant environmental interrelationship within the region the model begins to become increasingly unmanageable.

Isard and colleagues readily admit that the systems of linkages between the natural and social environments are extremely complex. Because the total system is composed of intricately interrelated variables, few, if any, of these linkages can be accurately represented in simple model form. For this reason many researchers avoid empirical investigations because such an approach cannot possibly portray accurately and comprehensively the interdependent structure of the ecological and economic systems. However, as

Isard states in his book, in many situations critical decisions must be made. Water and air pollution must be controlled now: we cannot afford to postpone decisions until models with a high potential for accurate prediction have been developed.

Despite such limitations, the Isard study is an important contribution to the analysis of economic-ecologic linkages. This model has provided the theoretical foundations and framework from which numerous later studies have been based. These studies propose models which are more operational than that of Isard's by eliminating consideration of the economic-ecologic interrelationships.

The following sections describe the mathematical formulation of the more operational ecologic-economic linkages models. Following this general discussion is a brief summary of several of the models which have been developed as a result of the Isard study.

V. 3. General Description of the Modified Economic-Ecologic Linkages Models

The economic-ecologic linkages models are modifications of the standard interindustry I-O transaction model, and as such, have the usual data requirements of inputs from industry i to industry j. Thus, these models produce the traditional estimates of direct, and indirect production in the region due to expected changes in final demand. By including into the transaction matrix employment coefficients and the households sector estimations can be made respectively, of total employment, total income generated, and induced production

due to estimated final demand. In light of the earlier discussion of the I-O technique, these are simply the traditional products of an interindustry I-O analysis.

V. 3. A. Estimation of Total Resourse Useage/Pollutant Emissions

Modification of the traditional model comes in the form of the generation of resource input and pollutant output coefficients for each sector within the model. Resource coefficients represent the amount of a resource (e.g. water in gallons, land in ft^2) industry j uses per dollar of j's output. Similarly, the pollutant coefficients represent the amount of a particular pollutant (e.g. SO_2, suspended solids, minerals, etc.) emitted by industry j per dollar of j's output. Premultiplying the (A) and $(I-A)^{-1}$ matrices by the matrix of resource and pollutant coefficients results in respective matrices of the direct, and direct and indirect resource requirements and emitted pollutants per dollar of final demand for each sector within the system. These resulting matrices enable the analyst to identify those industries consuming large amounts of natural resources and emitting significant amounts of pollutants during the production of their goods and services. By post multiplying the resulting matrix of direct and indirect coefficients by a column vector of final demand, estimates may be made of the total amounts of resources comsumed and pollutants emitted due to changes in final demand. In notation, the matrix multiplication can be represented:

$$(G) \times (I-A)^{-1} = (R)$$

$$(R) \times (V) = (T)$$

where: (G) = matrix of resource and pollutant coefficients

$(I-A)^{-1}$ = Leontief interindustry inverse matrix

(R) = matrix of direct and indirect resource and pollutant coefficients

(V) = column vector of estimated final demand

(T) = matrix of total output of pollutants - total resource input resulting from changes in final demand.

To illustrate the above matrix manipulations consider a three sector economy whose $(I-A)^{-1}$ matrix is shown in Figure 5.

			sectors 1	2	3
	agriculture	1	1	2	3
sectors	chemical	2	1	3	3
	steel	3	1	2	4

Figure 5 $(I-A)^{-1}$ matrix

Assume that each consumes one environmental input (land) and each produces one environmental output (particulates) as given in Figure 6.

	1	2	3
land use (acres)	2	2	3
particulates (#'s/day)	1	3	2

Figure 6 G matrix

The direct plus indirect resource and pollutant coefficients are determined by:

$$\underset{(G)}{\begin{vmatrix} 2 & 2 & 3 \\ 1 & 3 & 2 \end{vmatrix}} \times \underset{(I-A)^{-1}}{\begin{vmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{vmatrix}} = \underset{(R)}{\begin{vmatrix} 1 & 2 & 3 \\ 7 & 16 & 24 \\ 6 & 15 & 20 \end{vmatrix}}$$

If final demand increases by the amount given in the (V) matrix below, the total increase in land useage and particulate emissions may be found by:

$$\begin{vmatrix} 7 & 16 & 24 \\ 6 & 15 & 20 \end{vmatrix} \quad (R) \quad \times \quad \begin{vmatrix} 10 \\ 5 \\ 7 \end{vmatrix} \quad (V) \quad = \quad \begin{vmatrix} 318 \\ 275 \end{vmatrix} \quad (T)$$

land (acres)
particulates (#'s/day)

Thus conceived within the I-O framework, the ecologic materials flow can be seen as a special type of sales and purchase, i.e. inputs and outputs, or interdependence between the economic and ecologic systems. Comparison of the estimated (T) matrix with the estimated economic benefits gained from increased output will give some indication of the true environmental costs at which such benefits are gained. Or quite simply, it provides an estimate of the total pollution resulting from expected increases in regional output. Uses of the model and the estimates which it generates will be discussed later in this report. For simplicity, further discussion will be limited to the pollutant coefficients and the estimates that result from them. The conclusions and examples given are equally valid for the resource usage coefficients and their resulting estimates.

V. 3. B. Pollutant-Income Estimates

In many instances total increase in income, not output, is deemed the most relevant consideration in a regional planning analysis. Estimates of tons of steel or pounds of nails produced

are often difficult for the planner, citizen or public official to relate to in terms of economic meaning. On the other hand, estimates of the total income generated within the area due to increased output are easily grasped by all. Thus it is often useful to relate the estimated levels of emitted pollutants to income. This relationship is obtained by dividing the (R) matrix by a diagonalized matrix of income multipliers obtained from the interindustry model (households endogenous). A resulting coefficient is the waste load of pollutant i resulting from a one dollar change in local income arising from increased output in sector j. This may be represented as:

$$(R) \times (\hat{M})^{-1} = (E)$$

where: (R) = direct and indirect pollutant coefficient matrix

$(\hat{M})^{-1}$ = inverse of a diagonalized vector of income multiplier

(E) = environmental - income multipliers.

The resulting multipliers are approximations of the trade-offs between increased income due to additional output in sector j and the increased amount of emitted pollutants resulting from such output increase.

V. 3. C. Pollutant - Employment Estimates

A similar economic-ecologic relationship can be estimated using the employment multipliers. Because proposals to expand an existing industry or to attract a new industry to the region are often promoted in terms of the resulting increase in jobs, an estimation of the trade-offs in terms of employment is necessary to accurately

evaluate such proposals. Again the (R) is divided by a diagonalized matrix of multipliers. These employment multipliers represent the employment requirements for sector j per dollar of j's output. Thus a coefficient in the final (F) matrix represents the increase in pollutants resulting from a hiring increase of (x) employees due to increases in output in sector j.

$$(R) \times (\hat{N})^{-1} = (F)$$

where: (R) = direct and indirect pollutant coefficients matrix

$(\hat{N})^{-1}$ = inverse of a diagonalized vector of employment multipliers

(F) = environmental - employment multipliers.

V. 4. Historical Development of the Model

The above description of the model is actually a conglomerate of applications and modifications by several different authors. Each report, summerized below, represents an improvement over previous studies. While the summaries are brief, the major technique employed by the particular author is described.

V. 4. A. Environmental Planning: An Economic Analysis (Hite & Laurent, 1972)

The Hite and Laurent study confronts the delicate problem of effectively allocating resources between pecuniary uses and environmental preservation. Like the Isard model, the Hite model attempts to identify economic-ecologic linkages, and determine the environmental impact of the activity of various economic sectors within a region. Although the study was conducted for the purpose of analyzing coastal zone resource management, the conclusions and

implications drawn from the technique are generally applicable to the study of any area.

As mentioned in the previous section, many of the operational problems of the Isard model stem from its all-emcompassing nature. The present model under discussion is a modified version of the Isard model. Ecological inputs into the economic system can be viewed at many different levels of aggregation. Inputs may be classified as beach, marsh, and fast ground, or all such inputs can be lumped together simply as "land", the level of aggregation necessary, being determined by the particular problem under study. When one attempts to construct the universal model proposed by Isard, one will be forced to go to considerable detail in specifying exports from the ecologic and economic sectors because of the enormous complexity of the ecological interprocesses. However, achievement of the objective of identifying economic-ecological linkages does not require such a comprehensive model. In their study, Hite and Laurent were interested only in the junctures of the economic and ecologic systems, leaving the complex ecologic interprocesses to other, more competent professionals who study these processes.

Figure 2 graphically illustrates the modified version of the Isard model employed by Hite and Laurent. Except for the exclusion of the ecologic interprocess matrix in the right-hand corner, the model is identical to the Isard version.

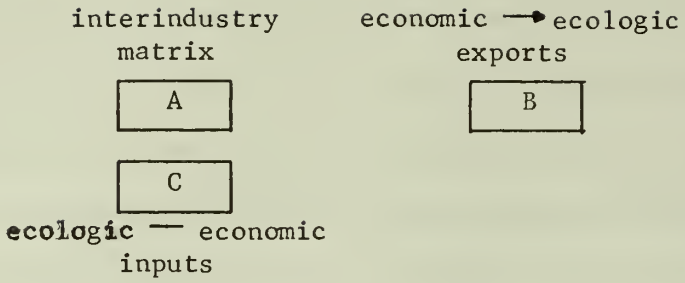


Figure 2

where: A = standard interindustry I-O matrix

B = matrix of economic outputs (pollutants) to the environment

C = ecologic inputs into the economic system.

The operational advantages of the modified model relate to the size of the B and C matrices. This type of model contains N number of industrial sectors in the A matrix, M number of ecologic-economic imports in the C matrix, and K number of economic-ecologic exports in the B matrix. This finite set of numbers is not constrained or expanded by Isard's ecologic interprocess matrix. By removing the severe constraint of obtaining data describing the ecologic interprocesses, a more operational and manageable model is obtained. Also, this exclusion detracts little, if any, from the ability of the model to estimate ecologic impacts resulting from economic activity.

In actual use, the (B) and (C) matrices are combined into one matrix (R), representing resource (ecologic) inputs and pollutant outputs per dollar of industry j's output. Then, as described in Section III. 2. A., the environmental matrix (R) is "mated" to the (A) and (I-A) matrices to obtain estimates of the direct, and

direct, indirect and induced (households endogenous) environmental coefficients. This resulting matrix of coefficients describes the total resource and pollutant impacts for each sector due to a one dollar delivery to final demand. Application of these coefficients to estimated changes in final demand provides an estimate of the total resource usage and pollutant emissions.

Also included in the Hite and Laurent study is a method for attempting to relate environmental quality to a pecuniary value. This technique is based upon the assumption that the demand price of environmental goods (environmental quality) can be estimated from an estimation of society's "willingness-to-pay" for such goods. Such willingness-to-pay estimates may be obtained from proxy measures such as land values or household surveys. If reliable estimates of the amount that society is willing to pay for various levels of environmental quality can be obtained, a rough value for the demand price of such quality levels can be estimated. This demand price can then be compared to the generated income estimates obtained from the I-O matrix. Suppose increased output from industry j to meet final demand results in a pollutant emission that is above that deemed desirable. The amount that society would be willing to pay to treat this pollution to reach a given level of quality must then be compared with the amount of income generated to the community by that level of j's economic output. If income generated is less than the willingness-to-pay value,

it would benefit society as a whose to prohibit such economic output. This statement however must be considered in terms of the equity of such a prohibition. Questions such as who (which groups) pays for the cleanup, who suffers from the pollution, and who receives the income that is generated, must first be answered before a restriction of economic output could be justified.

V. 4. B. An Environmental Linkages Model of the State of Maryland (Korbach, 1972)

The environmmntal linkages model of the state of Maryland is based upon the concept of materials balance and the conservation of matter, that is, matter is neither created nor destroyed. Thus, inputs of raw materials and other finished goods, which are used in the production of goods and services, ultimately result in waste residuals through either processing, delivery, or final consumption. This study is concerned with tracing materials flows from the initial generation of wastes, through the various routes to their ultimate disposal. For a regional economy with extensive production and environmental flows, the transfer of wastes must be identified. The Maryland model considers the problem of total emissions, leaving analysis of the physical movement of these emissions to the numerous diffusion models which are in existence.

Because of the complexity of the materials flow process a comprehensive accounting system for waste flows is necessary. Without a systematic approach to the investigation of emissions and wastes, double counting of emissions will likely occur. The Maryland

model employs such an accounting procedure for tracing waste flows from their production origin to their ultimate disposal in the environment for any given economic sector.

The initial (gross residual) waste generated by a firm is considered a function (P) of its total output. (P) represents a matrix of coefficients of emitted pollutants per dollar of output (for all sectors within the model). Gross residual may be discharged untreated by the firm (b_1), treated by the firm (b_2), or contracted to another firm for untreated disposal (b_3). Gross residual treated by the firm may then be handled in three possible ways: discharged to final receptor medium (d_1), recycled into the production process (d_2), or contracted to another firm for disposal (d_3).

Through the use of personal interviews, national studies, engineering texts, and unpublished reports, estimates of the 7 coefficient matrices, (P), (b_1), (b_2), (b_3), (d_1), (d_2), (d_3) have been made. Through the use of the detailed waste accounting model (92 industries, 49 waste materials), the policy options for environmental improvement are readily identified and easily measured. The most obvious option would be to possibly alter the percentages of gross residuals going to the treated and untreated processes before ultimate disposal. The coefficient matrices enable these initial amounts along with the effects of alternative amounts of treatment to be estimated. Other options available are the use of either cleaner inputs (decreasing gross residual), or of more efficient

production processes.

Utilizing a matrix of gross residual coefficients in combination with the interindustry model, three additional measurements of waste loadings may be derived. As described in Section III. 2. A., estimates of the direct and indirect wastes generated per dollar of final demand may be obtained by premultiplying the Leontief inverse by a matrix of pollutant coefficients (gross residual - P). Similarly, as discussed in sections III. 2. B. and III. 2. C., estimates may be made of the pollution resulting per dollar of income and employment position generated by changes in final demand.

The Maryland model and the estimates which it produces reveals numerous options for reducing pollution and improving environmental quality. The more obvious options are to treat untreated wastes, to recycle wastes, to use cleaner inputs, to improve technology, to modify the location of emissions, and to modify the total magnitude and composition of final demand (and thus, gross output). Before a decision can be made as to which policies to choose and how far to apply them, the equity or pareto-optimality aspects of such options must be investigated and identified.

V. 4. C. Air Pollution Abatement and Regional Economic Development (Miernyk and Sears, 1974)

A dynamic I-O model has been recently used by Miernyk and Sears to estimate the total costs of achieving national air quality standards in the state of West Virginia. This technique represents still another innovative approach to analyzing the linkages between the economic and ecologic systems, as well as introducing some

new innovations into I-O modelling at the regional level.

V. 4. C. a. General Description

The annual output required to meet estimated changes in final demand can be estimated by solving:

$$X_t = (I - A_t - D - B)^{-1} (Y_t - BX_{t-1})$$

where: X_t = total annual output for year t

A_t = technical coefficient matrix for year t

D = replacement capital coefficient matrix

B = expansion capital coefficient matrix

Y_t = final demand vector for year t.

The West Virginia model has been used to project the output of its 48 sectors to 1975. It also provides the information needed to estimate the investment required to achieve these output levels. Since a matrix of labor coefficients is available, it is possible to estimate the employment by sector that will be associated with projected output levels. All projections were first made before air pollution abatement standards were adopted. Thus they represent projected output levels that are not constrained in any way by the cost of pollution abatement.

Estimation of the costs of achieving given air quality standards requires detailed knowledge of the input structure if the anti-pollution processes. This included data on the installation and operating costs of pollution abatement equipment for the major industrial air polluters located in West Virginia. The types of data to be collected are discussed later in this report. The objective

of the Miernyk study was to collect primary data that would permit an adjustment of the elements of the basic matrices in the I-O model to reflect the added costs incurred by the region's major polluters as they comply with prescribed emission standards. Estimation of the direct and indirect costs of compliance were obtained by solving the following:

$$X_t' = (I - A_t' - D' - B') (Y_t - B'X_{t-1})$$

where the primes indicate that some elements in each of the matrices have been changed to reflect the capital and operating costs of pollution abatement. The final "real" costs incurred are then obtained by:

$$X_t^* = (X_t' - Y_t) - (X_t - Y_t)$$

Because a constant level of final demand was specified for both output estimates, it is necessary to subtract this constant term from both projections to accurately measure the relative change in interindustry costs due to pollution abatement. Thus, X_t^* yields the estimated incremental production that will be required to meet a fixed amount of final demand when certain establishments in the state have complied with existing pollution abatement requirements. The elements of this vector show the distribution of the additional production requirements by sector.

V. 4. C. b. Pollution Coefficients

Before an analysis of the costs of pollution control technology is possible, some estimate of the types and amounts of pollution must obviously be obtained. Similar to the other linkages models discussed,

this involved the estimation of the direct and indirect air pollution coefficients for each of the relevent sectors in the model. Three types of pollution coefficients were generated; particulates, sulfur dioxide, and hydrocarbons and carbon monoxide, although control technology for only the first two was discussed in the report by Miernyk. Although not explicitly stated, it is assumed that the coefficients were derived from various sources including survey data, engineering texts and national studies.

V. 4. C. c. Changes in the Coefficient Matrices

Several alternative systems of control technology were examined for particulate and sulfur dioxide pollution abatement. Some of those considered included; baghouse filters, wet scrubbers, coal-cleaning, and flue-gas scrubbing. Although the Miernyk report included separate chapters discussing the costs involved for control of both types of pollutants, only the total costs, i.e. combined costs for controlling both, will be discussed here.

As mentioned earlier, the total costs of air pollution abatement is estimated to be the increase in production required of each sector to meet a specified level of final demand. This increase is the result of changes in some of the coefficients of the I-O matrices. Changes in these coefficients are due to several factors. First there are the capital costs of purchasing and installing the pollution control devices. These costs by far represent the largest total dollar outlay, and thus it is the capital coefficients which

experience the largest change. In addition, there are operating and maintenance costs associated with each alternative control system under consideration. Finally, addition of control technology will result in changes in the input structure of those industries installing the equipment. This includes such factors as additional electricity, water, chemicals, labor, plant expansion and possible modification of existing production equipment required per unit of output. Examination of the input coefficients reveals that not only is the total output of the major polluting industries going to be affected by emission control, but also the output of the small, "non-polluting" sectors which supply inputs to those industries. Thus for example, it was determined for the West Virginia economy that 16 industries are direct sources of particulate pollution and only 3 are emitters of sulfur oxides. But because of the interdependence on both the input and output sides, all sectors of the economy contribute indirectly to air pollution.

V. 4. C. d. The Total Capital Costs of Clean Air

Some economic activities in West Virginia will experiance economic gains from the installation and operation of pollution abatement equipment. This is particularly true of general building contractors and of banks and other financial institutions that will help finance the large capital outlays that will be required. But the total changes in output may be considered as accurate reflections of the cost of clean air since they are the changes in production that would be required to produce a constant vector of final demand.

While it is impossible to fully assess the implications of the total costs associated with alternative levels of treatment technology without actual knowledge of the various alternatives and the West Virginia economy, examination of the absolute values will give some indication of the types of impacts estimated and the information provided by the model.

Total capital investment required by alternative pollution abatement programs ranged from slightly more than $352 million to almost $976 million. This amounts to a total capital cost of from 2.5 to 7.2% of the estimated Gross State Product. Annual operating costs ranged form $55 million (0.4% of GSP) to $152 million (1.1% of GSP). The above cost estimates relate only to pollution generated by stationary sources, with no account being taken of the large contribution that automobiles and other possible sources make to total air pollution. Furthermore, these figures describe only the direct costs of pollution abatement. It is necessary to examine the direct plus indirect costs to fully evaluate the total impact of industrial compliance with West Virginia clean air standards.

V. 4. C. e. The Direct and Indirect Costs of Clean Air

As stated earlier, the total changes in output required to meet a given level of final demand may be considered as accurate reflections of pollution control costs. Again looking at absolute values, total interindustry transactions, (due to alternative levels and types of emission control) would increase anywhere from $144

million (2.18%) to $496 million (9.7%), depending upon the alternative chosen. Imports would increase by $163 million (2.7%) to $309 million (5.1%). Payments to households would increase by $13 million (0.2%) to $37 million (0.6%). Total gross output to produce a constant volume of goods and services for delivery to final demand would increase anywhere from $323 million (1.6%) to $864 million (4.3%). There are substantial sector-to-sector variations in the various alternative approaches considered. However, all sectors obviously experience increasing costs (as measured by changes in sectional output) due to attempted compliance with air quality standards.

One must now ask what are the economic consequences associated with such pollution control costs? Certainly cost increases of the magnitude discussed cannot or will not be absorbed by industrial (and commercial) establishments. These increases will be passed on to consumers in the form of higher prices. These cost increases, however, would not be expected to place any particular area or region at a competitive disadvantage. Because the clean air legislation is national in scope, manufacturing industry would be required to comply with established standards wherever it is located.

V. 4. C. f. Conclusions

The Miernyk approach adds yet another potentially valuable tool for examining economic-ecologic linkages. The innovation of this technique, obviously, is the use of an I-O model for not only estimating the amount of pollution associated with economic output,

but also for estimating the various impacts or costs of controlling that pollution.

Because the largest cost increases are due to capital expenditures, use of the technique would require a matrix of capital requirement coefficients. At the very minimum, it would require the addition of a capital row and column to the I-O model. However this approach would only yield an estimate of the total capital costs incurred, with no information on which sectors would be experiencing what amounts of that increase. Obviously, examination of the pollution generated in a region, and of the various control technology available, would give some indication of which industries would be required to install new equipment. However, this would not provide information on the indirect costs of producing and installing that equipment. That is, the indirect sectoral linkages to the major polluters would be masked.

The Miernyk approach possesses great potential in that it can be used to analyze various kinds and levels of emission standards (e.g. air, water, primary, secondary). Any agency concerned with the establishment and implementation of environmental quality standards must have some idea of the economic consequences associated with those standards. With the Miernyk approach, those industries which will experience the largest cost increases due to pollution control standards can be identified and the costs estimated. The technique's greatest contribution is that it presents a means of evaluating the

tradeoffs between achieving alternative levels of environmental quality and the economic costs associated with such quality.

V. 5. Limitations and Resource Requirements of the Economic-Ecologic Linkages Models

V. 5. A. Data

The enormous amount of data required for the interindustry I-O model has been discussed earlier in this report. Obviously, since the interindustry matrix is a vital part of the linkages model, this requirement also applies to a model which attempts to identify and analyze economic-ecologic interrelationships. However, the data problem is further compounded by the fact that such models also require the derivation of environmental coefficients, i.e. resource inputs and pollutant outputs. As the economic and pollutant diversity in a region increases as would the amount of data required and the cost of obtaining such data. Because information on the interindustry data collection process is contained in this and numerous other published reports, the following discussion on data sources and limitations is limited to that required for estimation of the environmental coefficients. As in Section V. 2., which describes the linkages model, the following discussion is a composite of the three economic-ecologic studies described earlier in this report and detailed, for each technique, in the appendix.

(1) Sources: As mentioned earlier, data for estimation of the environmental coefficients can be obtained through personal interviews and surveys, national studies, natural science reports,

engineering texts, and numerous unpublished reports. The personal interview and survey technique has the distinct disadvantage that data from many firms is unavailable, either because the firms do not know the amount of pollutant emissions, or are simply unwilling to disclose such information. This would possibly suggest that national studies, where the data are more likely to be obtained, should be the primary data source. However, use of national data, as is also the case for interindustry coefficients, for estimating environmental coefficients assumes that the national product mix and national technology is identical to the product mix and technology of the study region. Certainly some industrial standardization exists, i.e. steel production through a hot-rolling process in Indiana will likely produce emissions similar to those from a hot-rolling plant in Pennsylvania, assuming identical inputs. For numerous other activities, however, the similarity most likely does not exist. Thus, as the economic activities and processes in a region deviate from the national averages, the probability of national coefficients accurately describing regional relationships decreases rapidly. For those areas containing new industries, one would expect more efficient, up-to-date technology than in those areas containing older, established firms. Likewise, with the more advanced technology, one would further expect less pollutant output than that of the older firms. Data obtained through engineering texts are often not in a form suitable for inclusion into the model.

For example, information on air pollutant emissions may be based upon physical units rather than dollar output. Consequently, to use these data within the format of the model, it is necessary to assume prices for the physical units of capital. Similarly, data may be reported on a "per employee" basis, rather than a dollar output or physical output basis, again requiring some type of conversion. Similar problems may also exist when data are gathered from other technical - natural science reports. This all suggests that at the present time the most suitable sources are the national studies which have been conducted for the purpose of identifying specific amounts of pollutant emissions. While the data are averages, they are at least a starting point from which preliminary estimates and guidelines may be generated. As one begins to construct an environmental coefficient matrix, one quickly realizes that given the present state of environmental science, much data are unavailable and serious gaps in our knowledge do indeed exist. However, identification of such gaps will provide direction for future areas of needed research. As such research is conducted and data acquired, we will begin to obtain a better understanding of the economic-ecologic linkages, and thus be better able to estimate their impacts upon one another.

V. 5. B. Assumption of Linearity

The usual criticisms of linear coefficients as applied to interindustry I-O analysis are applicable to the environmental coefficients. That is, we assume that the amount of pollutant

emitted per dollar of output is the same for either $500 of output or $50,000 of output. Undoubtedly, this assumption of linearity with regard to environmental linkages is unrealistic, and in some cases, may result in serious inaccuracies. Technology, as mentioned earlier, can change over time. For any particular industry the average coefficient may overestimate the waste load if it possesses new technology, or underestimate the load for a plant with older technology. One possible attempt at overcoming this problem is to identify the "best practice" firms, i.e. those with the most advanced technology, and assume that their input requirements and pollutant outputs will eventually represent the "average" requirements and outputs at some future point in time. In light of past industrial advancement and the future imposition of emission standards, this does not appear too unreasonable an assumption. Similar to technology mix criticisms are level of aggregation and product mix criticisms. For example, the average BOD output for the entire food processing sector may be quite different from the average BOD output of the dairy products component of that sector. For any particular sector the amount of emitted pollutants is a function of what products are produced and what inputs are used in that sector. Thus, as the amounts and types of products produced by a sector change over time, so does the amount of actual emitted pollutants. The linearity assumption criticisms increase when consideration goes beyond the nanufacturing sectors. For

example, in the construction sector, output may be doubled by building a twenty rather than ten story office, with very little additional soil erosion. Similarly in agriculture, a farmer may have to double the use of fertilizers and pesticides to gain only a 25% increase in output. The determinants of the amount of wastes generated in this case is not output, but rather soil type, slope, rainfall and cultivation patterns. Therefore, it can be seen that the pollutant projections are by no means absolute, but rather are only tentative estimates which must be interpreted as such.

V. 5. C. Pre- and Post-Treatment Waste Coefficients

The pollutant coefficients in the environmental matrix must be identified as either "before treatment" or "after treatment". If they are "before treatment" coefficients they represent raw output of wastes. If they are "after treatment" the level and type of treatment must be specified. This identification of treatment activity and efficiency obviously increases the complexity of obtaining data for use in the model. However national data sources do exist which provide estimates of treatment by type and efficiency. This type of coefficient disaggregation was employed in the Maryland environmental model and the reader is referred to the discussion of that model for a list of the data sources and advantages to such an approach.

V. 5. D. In-house Staff Expertise

Construction and use of a model as complex as the economic-ecologic linkages model will require the knowledge and input from

a number of professional fields. Ideally a staff should consist of a land-use planner, an economist-statistician, an ecologist, an environmental engineer, a geographer, and several subprofessional clerks and assistants. This suggestion is readily substantiated when one examines the types of data which must be gathered and adjusted for use in one of the complex linkages models. Such a staff could be assembled from in-house professionals, borrowed from other public agencies, contracted from private consultants, or some combination of the three. Obviously the size and composition of the required staff depends largely upon the size of the model, i.e. the degree of sector and pollutant disaggregation, and the purposes for which the study is conducted. For example, the extremely aggregated models, such as the TAP procedure discussed earlier, could possibly be constructed and operated by one or two highly qualified planner-economists and several assistants. This is the suggested minimun staff needed to supervise assembly of the needed planning tools and data, and to ensure their proper use.

V. 5. E. Time

The time required to complete an I-O analysis, like cost, is largely dependent upon the size of the model, its level of aggregation, and the data sources used. Because there was no specific mention of the exact time period required in any of the models reviewed, the author can only give an educated estimation. While the smaller aggregated studies, (e.g. TAP) could probably be

completed in approximately one year, the larger, more complex studies required roughly 4-5 years to complete, with the exception of the Maryland study which took slightly less than two years. This does not suggest that future studies will require as long for completion. Much of the time required for the described studies was spent in "breaking new ground", exploring new alternative procedures. By avoiding much of this required experimentation, and by incorporating already compiled data, completion of future studies should be possible within a much shorter period of time.

V. 5. F. Cost

Undoubtedly the major factor prohibiting an area from conducting an I-O analysis is the enormous cost involved with such a procedure. The cost of such an analysis, as expected, varies with the size of the model and whether the data is gathered through primary or secondary sources. The cost of constructing a matrix from survey data is roughly $100,000 to $500,000 or more, prohibiting all but perhaps a few large agencies from undertaking such a task. There are, however, at least two ways in which the costs of obtaining the needed I-O matrices may be reduced: (1) simulate the state and regional I-O matrices by adjusting the national I-O model, and (2) use of the smaller, more highly aggregated models such as TAP, which in certain cases can supply satisfactory results.

V. 5. G. Accuracy

Obviously, the accuracy of the estimates which a model produces

is extremely dependant upon the reasonableness of the theory and data upon which the model is based. In light of its numerous assumptions such as linearity and constant coefficients, the interindustry I-O technique has become generally accepted as a valid tool for making reasonable estimates of future economic impacts within an area. If one accepts this premise, then the economic-ecologic linkages model, which is simply an extension of the I-O technique, can also be labeled as a valid estimation technique. One would expect the accuracy of the linkages madel, however, to be somewhat less than that of an interindustry model alone, due to the fact that an additional source of possible error will exist by inclusion of the environmental coefficients. Because the coefficients for the most part are based mainly upon national data, the probability of them exactly reflecting regional conditions is slight. But these coefficients do at least enable some preliminary estimates to be made. They do provide the researcher a base from which to begin analysis of the economic-ecologic interrelationships. As the model is further adapted and modified, and as new research provides higher quality data, the accuracy of the linkages model will continually increase. In addition, the model's accuracy can be further increased by incorporation of the "changing coefficient" approach of the dynamic I-O model. For the present, the estimations which the model provides must be described as tentative. However, this present limitation does not disparage the value

of the model as an important and long-needed tool for regional analysis.

V. 5. H. The Value of the Economic-Ecologic Linkages Models

As mentioned earlier, regional scientists, planners and public decision-makers are becoming increasingly aware that we can no longer operate under the premise that the benefits gained from economic growth and development will always outweigh the costs of such development. The local tax benefits from such development are often exceeded by the public expenditures required to provide public services to the new development. There now exists the growing realization economic growth often brings with it large environmental costs. The widening public participation in government decision making and the increased public concern about the quality of economic development have generated a need for analytic methods which are sufficiently sophisticated to capture all of the economic, spatial, and environmental aspects of a development strategy, These models should be flexible enough not only to test the benefits and costs of alternative development proposals, but also should be able to assist in the formulation of optimal development strategies.

One of the most serious and challenging problems facing society is that of reconciling the demands for economic development with the requirements of environmental preservation. At one extreme are the "preservationists", who contend that all resources and areas should be preserved in some natural state and used, if at all, only

for some limited "safe" activities. At the other extreme are the "exploiters", who insist that all resources be used for purposes of economic development, as measured by increases in gross output, per capita income, and the number of jobs. Obviously and realistically, the optimal solution lies somewhere between the two extremes. Planners and policy makers are caught in the midst between these two extreme positions, and, without some objective basis for evaluating the tradeoffs between the two extremes, are vulnerable to the pressures exerted by each group. The goal of achieving the optimal solution can be achieved only if government officials, planners, and the public have access to objective information on the environmental impacts created by alternative economic development.

Identification and estimates of these impacts are provided by the economic-ecologic linkages models. These models generate estimates of both the economic growth (output) of an area, and of the environmental pollution that accompanies that growth. Through the generation and examination of the environmental-output, environmental-income, and environmental-employment multipliers an objective analysis of the tradeoffs is possible. However, the tradeoff estimates provided by the multipliers do not, in and of themselves, determine optimal policy solutions. These estimates are only a "supply price" analysis, that is, they can only describe the amounts of pollutant which are incurred per dollar of income or new job created. At present, we do not know the "demand price", the price society is willing to pay to clean up or avoid such pollution of

environmental quality. Environmental goods in economic terms are classified as "public goods", that is, a goods or services which if provided to and consumed by one individual, are consumed by all individuals. Because of the "public goods" nature of environmental quality, its dollar value cannot be determined through the market system. Without a knowledge of the price that society is willing to pay for various levels of environmental quality, the tradeoffs between pollutant increase and income increase cannot be objectively evaluated. When an estimate of the dollar amount which society will pay to avoid the increased amount of pollution generated by a one dollar increase in income is known, the total costs and benefits of such a production increase can be determined. If the amount that society will spend to avoid or clean up the pollution is greater than the income generated within the area, the increased production should be carefully examined and possibly prohibited.

One final comment as to the value of the impact estimates provided by the linkages model must be made. The estimates provide a framework for the analysis of the overall economic effects of various types and levels of emission standards which might be required of industries within a region. Comparison of the projected amounts of pollutants emitted by each industry with the maximum level or type of pollutant emissions permitted by local, state, or national standards provides an estimate of which industries will be able to install treatment processes, will exceed the emission standards by either expanding or locating within the region. This

type of analysis has important implications for those public agencies concerned with determining such standards to maintain and preserve environmental quality. Such an analysis, however, should not be regarded as an additional analytic burden. On the contrary, this type of economic analysis will not only air in determining optimal standards, but will also provide a basis of credibility upon which such standards can be justified and defended.

As has been discussed throughout this report, the techniques of and the estimates provided by economic-ecologic analysis are not definitive. No economic model, no model of any sort, can automatically solve the complex problems of effective environmental management and economic development. Further multi-disciplinary research is badly needed. While the present models do provide a foundation for examination of the economic-ecologic linkages, continual development and refinement of the linkages model is necessary. As we continually increase our knowledge and understanding of the interrelationships between the economic and ecologic systems, we will advance closer to achievement of the goal of optimal balance between these two systems.

Much of the validity of the I-O projections, whether they be of the more conventional or ecologic-economic type, is determined by the accuracy of the final demand estimates upon which they are based. The following section describes one technique for making accurate estimates of changes in final demand.

VI. Econometric Models

As is evident from the preceeding sections, the economic base and input-output techniques cannot be used for forecasting until projections of either the basic or final demand sectors, respectively, are available. Obviously the validity of the two techniques' projections is largely related to the quality of the final demand estimates upon which they are based. One method useful in forecasting various components of final demand is econometric modeling. These models represent a compromise between economic base and I-O by providing far more information than the base approach at a substantially lower cost than the I-O technique. In addition, econometric techniques may be used to test various economic structure theories, while economic base and I-O, on the other hand, are tied to specific theories. Thus econometric models represent a more flexible approach to the study of regional economies.

The following discussion is intended to provide the reader a cursory description of the econometric modeling technique. For ease of explanation, the description given is a very simplified explanation, and by no means covers the numerous components or the detail of such components which may be included in an econometric model. For a more comprehensive discussion the reader is referred to any of a number of excellent texts on the subject.

VI. 1. Study Area

Numerous large scale econometric models have been constructed for the United States, and the technique has become a rather

conventioaal approach to the study of national economies, exhibiting excellent forecasting qualities. However, use of the technique for subnational studies such as state and regional analysis, was for a long time either unattempted or done in a relatively simple fashion due to regional data collection constraints. Within the past few years increased interest in economic analysis at the state and regional levels has resulted in several valuable applications of the procedure at the subnational level. Although the following discussion of the general econometric model applies to the national studies, the description is conducted within a regional framework. For more explicit examples of actual regional econometric models the reader is referred to: "An Econometric Forecasting Model for the Philadelphia Region", N. Glickman, Journal of Regional Science, Vol. 11, No. 1, 1971, and "The Specification of Regional Econometric Models", L. Klein, Papers of the Regional Science Association, Vol. XXIII, 1969.

VI. 2. General Description of the Model

An econometric model consists of a set of n equations describing various economic components of the study area. Each endogenous variable (one determined by the model) is a function of other endogenous variables, exogenous variables (variables determined outside the model) and an error term. Each equation may be implicitly represented as:

$$Y_{it} = f(Y_{jt}, Z_{kt}, U_t)$$

where: Y_{it} = endogenous variable i in period t

Y_{jt} = endogenous variable j in period t

Z_{kt} = exogenous variable k in period t

U_t = error term in period t

A set of mathematical equations describing the structural interrelationships allows for simultaneous determination and projection of the local varibles. The determination of Y_{it} by a variable determined elsewhere in the model, Y_{jt}, is the essence of a simultaneous equation model.

The entire set of equations may be represented as:

$$\sum_{j=1}^{m} a_{ij} Y_j + \sum_{j=m+1}^{m+m} a_{ij} Z_j = U_i \qquad i+1,\ldots,m$$

Solution of the model involves determination of the values of the coefficients, a_{ij}. This is accomplished through the use of time-series (historical) data and some method of regression analysis, such as Ordinary Least Squares (OLS) or Two-Stage Least Squares (TSLS).

An example will serve to demonstrate the structure of an econometric model. Suppose it is desired to project the growth of a simple economy composed of only two goods, consumption and investment. Hypothesizing that the volume of consumption goods is primarily determined by the total volume of goods, the following equations can be formulated:

$$Y = C + I \qquad (1)$$

$$C = A + BY + U \qquad (2)$$

Equation (1) states that the total dollar volume of output (Y) equals the dollar volume of consumption goods output (C) plus the dollar volume of investment goods output (I). Equation (2) describes the variance of the level of consumption goods with the level of total output. The residual term (U) reflects influences on (C) other than (Y). The variables (C) and (Y) are the endogenous dependant variables and (I) represents a predetermined variable whose behavior is not explained by the model. Equations (1) and (2) are said to be in structural form because they are stated in the term that reflects the structure of the economy. Using historical data, the parameters (A) and (B) are determined by any of several methods such as (OLS) or (TSLS). Substituting the estimated parameters (A') and (B'), the model now consists of the equations:

$$Y = C + I \qquad (3)$$

$$C = A' + B'Y \qquad (4)$$

Although the model is now solved, projections cannot yet be made. To do this, the model must be in reduced form, with only predetermined variables on the right hand side of each equation. Then existing and projected values of the exogenous variables can be substituted in, and the equations solved for values of the endogenous variables. In the case of this example, equations (3) and (4) must be solved in terms of (I). Therefore, the reduced form would be:

$$Y = A'/1-B' + \left[1/1-B'\right] I \qquad (5)$$

$$C = A'/1-B' + \left[B'/1-B'\right] I \qquad (6)$$

Now, given the value of (I) for a particular year, the values of (Y) and (C) for that year can be determined. Given projections of (I), projections of (Y) and (C) can also be made.

In actual application, the econometric model may consist of 20 or more equations relating output, income, employment and consumption within an economy. Given the values of the predetermined variables the equations are simultaneously solved to provide projections of the various endogenous variables. Because of this simultaneous process, the variables interact upon each other much in the way they do in the actual economy.

VI. 3. Accuracy

Once the model had been specified, solved, and projected, one must determine how accurately it describes the real world. Several pieces of information help reflect the accuracy of the model, although none of them is foolproof. The first approach is an examination of the structural coefficients. The coefficients, which describe the magnitude of and relationship between certain variables, should conform with economic theory. In the above example, the value of B' should lie between zero and one. If it was less than zero, consumption output would decrease as total output increased, which is unlikely. Further, if B' were greater than one, consumption would be greater than total output, which is again unlikely. If the model contains coefficients which appear unreasonable, the specification of the model should be examined and possibly changed, and the model solved again.

Another measure of the accuracy of the structural equations is the coefficient of multiple determination R^2. This statistic describes the variation in the dependent variable which is "explained" by the independent variables. The closer R^2 is to one, the better the independent variables predict the behavior of the dependent variable during the historical period.

Another statistic which measures the goodness-of-fit of each equation is the standard error of the estimate (S.E.E.). Similar to R^2, it measures the dispersion of the observations of the dependent variable around the predicting equation. Statistical theory states that 66% of the time the actual values of the variables will lie within a range $\pm$ the S.E.E. about the estimates values. The closer this dispersion is to zero, the more accurate the model.

The model can be modified and revised through the application of various statistical tests to obtain the most accurate form. However, one important point must be mentioned. These methods help determine only how accurately the model reflects past behavior of the variables in the model. They do not help determine the probable accuracy of the projections. Projection accuracy similar to the descriptive accuracy of the model obviously depends upon the degree to which future conditions resembles past behavior.

The most appropriate test of the accuracy of the projections is their reasonableness in view of past behavior. Through the model, "projections" may be made for past periods for which actual data

exists. Comparison of the projected values with actual values gives an indication of how accurately the model reflects existing conditions. This test will not reveal the accuracy of the long-range projections, but it does provide some indication as to the possible reliability of the model in forecasting future activity.

VI. 4. Regional Econometric Models

As mentioned earlier, most econometric modeling has been done on a large-scale national basis. This level, however, provides little if any information to the analyst concerned with examining the structure of a smaller, regional economy. The following discussion provides a general summary of the types of considerations involved in conducting a regional econometric study. Sources of information for this discussion are the two articles mentioned in Section VI. 1. which describe regional econometric models mentioned in Section IV. 1.

VI. 4. A. Economic Components

The first step in constructing a regional econometric model envolves separating the local economy into several sectors such as:

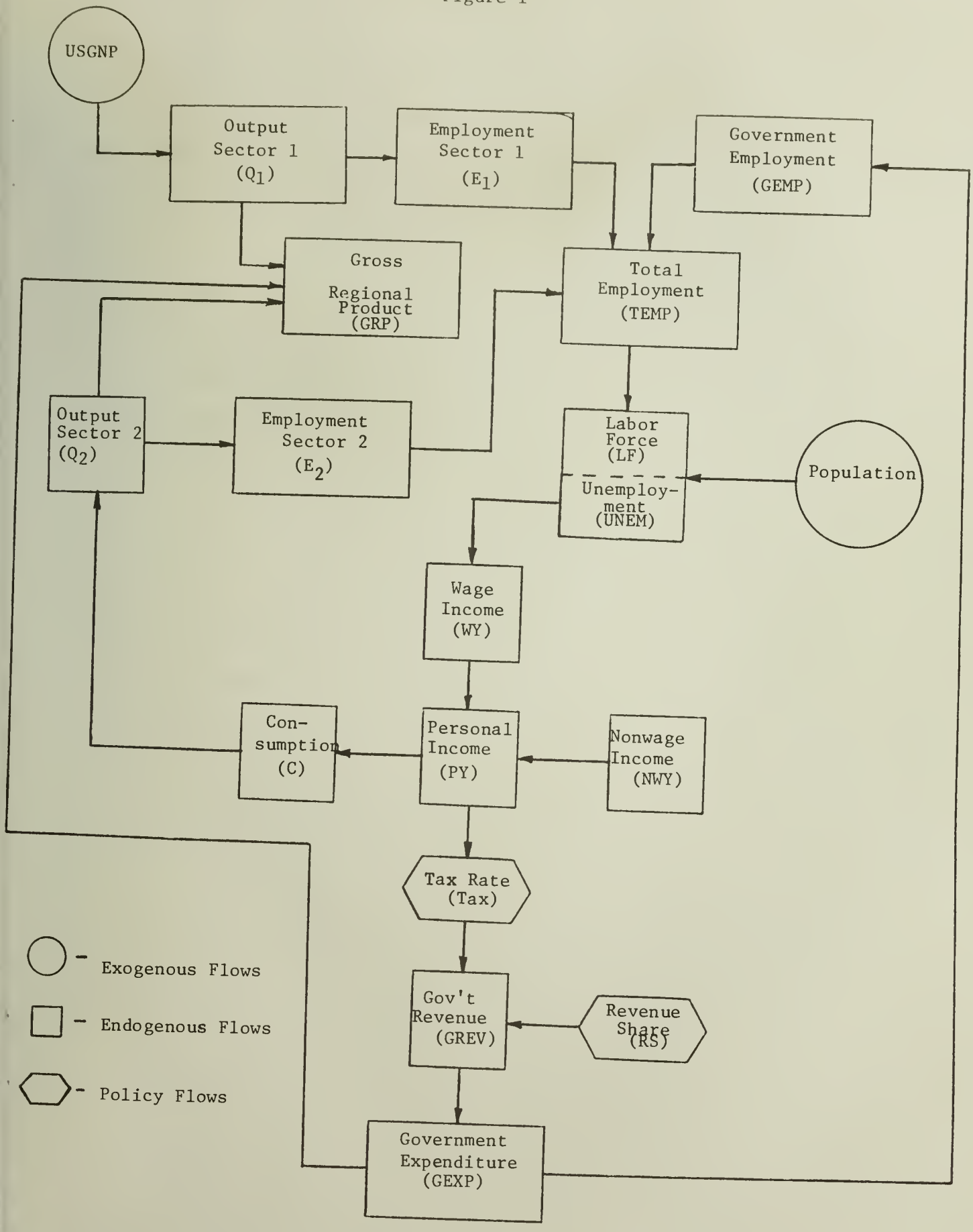
Sector 1: Manufacturing

Sector 2: Wholesale and retail trade

Sector 3: Government

With reference to Figure 1, the explicit embodiment of the economic base hypothesis can be seen in the assumption that output in sector 1,

Figure 1


USGNP
Output Sector 1 (Q_1)
Employment Sector 1 (E_1)
Government Employment (GEMP)
Gross Regional Product (GRP)
Total Employment (TEMP)
Output Sector 2 (Q_2)
Employment Sector 2 (E_2)
Labor Force (LF)
Unemployment (UNEM)
Population
Wage Income (WY)
Consumption (C)
Personal Income (PY)
Nonwage Income (NWY)
Tax Rate (Tax)
Gov't Revenue (GREV)
Revenue Share (RS)
Government Expenditure (GEXP)
- Exogenous Flows
- Endogenous Flows
- Policy Flows

(Q_1), is a function only of U.S. Gross National Product, (USGNP), i.e., manufacturing is base or export oriented with respect to the region itself, and thus is tied to the growth of the national economy. On the other hand, sector 2 is local-market oriented and its output is therefore shown to be dependent upon local income and spending. Employment in each sector and thus total employment are functions of output, and unemployment can be computed from the employment and labor force figures. Personal income and consumption are functions of wage and non-wage income. Gross regional product is defined as simply the sum of value added in each of the sectors. Lastly, personal income, the local tax rate, and revenue sharing yield estimations of local government revenues and expenditures, which are the basis for computing local government employment. This obviously is a simplified view of a regional economy, with the inclusion of other, more detailed components, such as additional output sectors, tax rate disaggregation, investment, and capital stock being possible.

After the theoretical structure of the economy has been determined, equations relating each variable flow are then formulated. The following equations are typical of the type that would be constructed for the simplified economy presented above. All coefficients are, of course, hypothetical.

$$Q_1 = 2547.03 + 14.2 \text{ USGNP} \qquad (7)$$

$$E_2 = 174.5 + 0.17\ Q_2 \qquad (8)$$

$$C = 2984.72 + 0.61 \text{ PY} \qquad (9)$$

Equations (7), (8), and (9) relate, respectively, output in Sector 1 to USGNP, Sector 2 employment to Sector 2 output, and consumption to personal income. As mentioned earlier, actual coefficients are determined by regression analysis employing historical data. This data must be obtained by either modifying national data, or using various other sources such as state employment records, state manufacturing and commerce records, etc., the types and sources of data obviously depending upon the components included in the model.

VI. 4. B. Estimation of the Model

Using any of a number of regression methods, either singularly or in combination, the variable coefficients are determined. Examination of the R^2 and other statistics provides some indication of the model's ability to reflect past conditions, and the significance of the various variables in the equations. Careful variable selection and structuring of the equations should result in regional models which possess very high R^2 values, thus "explaining" a large portion of the variance in the dependent variables. This procedure enables each estimation (regression) procedure to be evaluated as to which produces the most accurate coefficients.

An additional estimate of the models accuracy employed in the two regional studies was obtained by computing the mean average percent error (MAPE) for each endogenous variable, where:

$$MAPE_i = \frac{(Y_{i(t)} - Y^*_{i(t)})/Y_{it}}{N} 100$$

$Y_{i(t)}$ = actual value for the i variable at time (t)

$Y^*_{i(t)}$ = predicted value for the i variable at time (t)

N = sample size.

For both studies, with the exception of certain variables, the average MAPE value is approximately one to three percent. This high degree of accuracy is not uncommon and would be expected of other econometric models conducted on the regional scale.

VI. 4. C. Simulating a Regional Economy

Given the regression coefficients and the actual values of the exogenous variables for period (t) the model can then predict or simulate the actual workings of a regional economy. Estimates for period (t+1) employ lagged values (values for period (t) of the endogenous variables). In this manner the projections for a given year are based upon estimates for the preceeding year. Projections for a future year must of course begin with projected values of the exogenous variables, obtained from other projection techniques such as a national econometric model.

Another perspective on each model's accuracy was obtained by comparing actual and model generated results as functions of time. Both the Glickman and Hall regional models exhibited a close correlation between actual and predicted values. However certain discrepancies did exist. In the Hall study of the Los Angeles region, for example, the model was unable to detect and track the significant downturn in economic activity that occurred in 1970. Model predictions for that period, particularly output and employment in the manufacturing sector, remained at nearly their 1969 levels instead of the dramatically

lower observed values (approximately 10% below the previous peak). This result, of course, is not unexpected of a model that is based upon past averages in economic activity. Rather it merely points to the inability of any modeling technique to predict drastic, unexpected changes in economic activity. Similar over- or under-estimations during "quirk" periods occurred with the Glickman model. However for the other, more standard periods, both models approximated actual conditions quite well.

Since no regional data exist for explicit evaluation of model estimates for future years, these estimates can merely be interpreted in light of the reasonableness or stability of the model estimates for past years. As stated earlier, a high level of accuracy in estimating past conditions is no guarantee of the accuracy of future projections. However, if one accepts the assumption that future economic activity will generally follow past levels and trends, the model's projections can be accepted as being relatively accurate indications of future conditions.

VI. 4. D. Evaluation of Alternative Policy and Growth Conditions

By incorporating the proper variables such as variable tax rates into the model, the effects of alternative policies can be estimated. For example, in the structure diagramed in Figure 1, alternative tax rates can be substituted into the model's equations and the resultant effects upon government expenditure estimated.

Similar effects of alternative changes in the national economic policy may be incorporated into the model. Glickman (1971) examined

the effects of the following assumptions concerning national policy: (A) a cancellation of the business tax credit; (B) changes in the income tax surcharge from 10% to 5% and finally being discontinued; and (C) federal employee pay increases and additional Social Security benefits and contributions. These various assumptions were then incorporated into the model to determine their impact upon the regional economy. An advantage over either the economic base or I-O technique, the econometric model represents a flexible tool for examining the effects of alternative policy proposals upon a regional economy.

VI. 4. E. Impact Sensitivity and Regional Multipliers

Related to the above section is the concept of impact sensitivity. An important aspect of the behavioral pattern of an economy is the sensitivity of the economy to various external shocks. This sensitivity is conventionally summarized in an impact multiplier, which is defined as the first period change induced in an endogenous variable in the model by a one unit change in a particular exogenous variable. For linear models the impact multipliers are given by the coefficients of the exogenous variables in the reduced form equations. These may be computed numerically rather than analytically be calculating changes in the endogenous variables that are caused by small changes in a single exogenous variable. Calculation of such multipliers, for example, would enable changes in the output of the export sector due to a one billion dollar increase in U.S. Gross National Product to be estimated. This change in Q_1 can then

be used to estimate the dollar increase in gross regional product. Similarly multipliers, and thus changes in employment and income due to exogenous changes, may be estimated. Although the I-O approach is a disaggregated procedure for computing average multipliers, the aggregated econometric multiplier approach results in multiplier values which compare quite favorably with those obtained through the former technique.

VI. 4. F. Summary and Applicability of Regional Econometric Models

Regional econometric models which exhibit accuracy comparable to that of the national models have been and are being constructed for the analysis of regional economies. These models provide a reliable means of estimating the final demand sectors of an economy, thus representing a technique which can be used in conjunction with a regional I-O model. The Glickman and Hall studies have demonstrated both the feasibility of constructing a regional econometric model, and the reasonableness of the impact multipliers generated by the technique.

Although no estimates were given by either of the authors, the regional econometric model can be constructed in considerably shorter time and with a smaller expenditure of money than that required for an I-O model. Staff expertise would also be expected to be less than that required by the construction of a linkages model. Obviously this would require several professionsls trained in econometric-statistical theory.

The benefits of saved time and money are not without their own costs. As evidenced by the extreme aggregation of the econometric sectors, the model can only provide grossly aggregated estimates of output, consumption, employment, etc. To provide estimates for individual industries within the economy would require separate equations and data describing those industries. This appears to be the most obvious limitation of the present econometric models. There is certainly a current need for more explicit consideration of at least the more significant industries within a local economy. Because major industries represent possible contributions to a region's volatility, their separate examination seems warranted. Future models will need to incorporate such aspects into their structure. When compared to the detailed projections provided by the I-O model, the econometric model projections appear as a rather gross aggregation of information.

Nonetheless, the econometric model does represent a valuable tool for the analysis of a regional economy in that it does provide total demand estimates and a means for evaluating alternative policy proposals. Whether or not such a model should be incorporated for the analysis of various regions of the State of Illinois obviously depends upon the purpose of the analysis and the types of projections desired. If one is interested in only average projections of either the basic or non-basic sectors, total consumptions, employment, etc., the technique is highly recommended. Past performances and the

ability of the model to be continually revised and modified suggest that highly accurate estimates are possible. Detailed projections for individual industries, however, will require that either new equations and data be incorporated into the model and tested, or else some type of I-O approach must be used. When such detailed information is not required, the accuracy, and time and money savings would suggest at least a consideration of the incorporation of an econometric model into any attempt to analyze and project a regional economy.

VII. Land Use Allocation Models

VII. 1. Introduction

Land use allocation models are a class of models concerned with estimating the future locational patterns (land use) of various urban activity components, e.g., housing, industrial employment, commercial and retail establishments. These models therefore typically allocate projections of total population and employment growth of an area (state, county, town) to specific sites within that area. Such projections can be either exogenous (made outside the allocation model) or generated by the model itself.

The spatial organization of population and economic activities within regions and urban areas influences planning in several important aspects. It is related to the basic concept of the planning process in that most decisions depend in some degree upon assumptions that have been made about the factors governing the locational choice of households, institutions, or businesses. They must take account not only of the spatial relationships surrounding each decision-making unit but also the extent to which individual choices are restricted by competition from other activities.

Urban and regional planners are concerned mainly with two aspects of spatial organization. First, they are interested in the factors governing the patterns of location of particular types of activities within a limited area. Second, planners are concerned with the flows between different parts of the study area and the ways in which population responds or will respond to existing or future

transportation facilities.

VII. 2. The Determinant of Spatial Organization

Urban spatial organization is the outcome of a process which allocates activities to sites. In our society, the process is mainly one of transactions between owners of real estate and those who wish to rent or purchase space for their homes or businesses. The market process of transactions between willing buyers and willing sellers determines the spatial organization of urban activities to a large degree. There are obvious exceptions, such as governments exercising the power of eminent domain, to this rule. However, it is the problem of identification, quantification and simulation of the various factors affecting the selection of a site which the land use allocation models confront. Certainly, household sites are selected on the basis of proximity to employment, open space, households of similar income, amenities, etc. Likewise industrial and commercial sites are chosen for availability to supplies, market proximity, transportation and other public services, etc. These and numerous other factors play a significant role in the site selection process. Thus any model which attempts to approximate and predict such a process must certainly incorporate some if not all of these factors into its structure.

VII. 3. The Significance of Allocation Models to Air Quality Maintenance Planning

As stated in Volume 4 of the Guidelines for Air Quality Maintenance Planning and Analysis, sources of air pollution may be divided

into three generalized categories - area, line and point sources. This is obviously not the only way to categorize sources of pollution, but such a system does allow the utilization of several land use planning concepts. Drawing heavily from the suggestions of this volume, a brief explanation of the importance of land use allocation models to air quality planning is presented below.

(1) Area Sources - This category is, for the most part, comprised of residential land uses. Determination of the amounts of emissions generated by this category requires knowledge of the location of residential areas, the types and densities of residential structures, and the size and distribution of the population. While zoning ordinances provide information on the densities of residential zones within an area, it is the allocation model which can provide approximations of the location of future residential activity. Also included in this category are commercial and office facilities, both of which are included in the output of an allocation model. Consideration of the results of such models in conjunction with the local and regional land use plans will enable better estimation of the future locations of area sources of emissions.

(2) Point Sources - Single major emitters (industries) located at a point are the source of attention of this category. Like residential uses, there do exist allocation techniques for determining the future location of new or expanding industries. Although this type of modeling has several major flaws (discussed later), they

do enable rough approximations to be made. Likewise, zoning and land use schemes, which often distinguish between light and heavy industries, can and should be incorporated into the location analysis.

(3) Line Sources - This category denoted major highways or other transportation links. Transportation planning requires that estimates of the existing and future locations of population and economic activity be made. From this information, data on the number and distribution of existing and future trips can be generated. Thus the allocation model provides a very important input to the determination of the origin and destination of urban tripmakers. This information along with that on planned transportation facilities (3-C process) enable estimates of the total line source emissions to be made.

Thus the allocation modeling process does indeed play a significant role in effective air quality maintenance planning. Estimation of expected total population and employment growth for an area is not, in and of itself, sufficient. It is the future location of this growth which must be identified before any accurate estimation of the amount and effect of the emitted pollutants can be made.

VII. 4. The Types of Models Reviewed

While several allocation models were examined, those selected for discussion in this report are limited to the simpler, least costly techniques. There are several reasons for such a decision. First, a model is obviously of little use if it is not operational.

While there are several models which are in existence, they are often far from being operational for areas other than that for which it was developed. This will become clearer in light of the second factor: cost. The complex, large-scale allocation models are developed at a very high cost of both time and money. Even application of fully developed models will require considerable expenditures. Because of the numerous and diverse variables included in the structure of the complex models, much of the required data does not exist in the average agency's file. This necessitates the generation of data on such factors as development costs, location probabilities, etc. Third, there is the technical expertise required to operate and interpret the results of such models, again usually available on a very limited, if at all, basis.

VII. 5. Industrial Allocation

As stated in Section VII. 3., the location of industrial or "basic" employment forms the foundation upon which most, if not all, land use models allocate population (households) within an area. While other factors, such as public services, housing density, open space, and amenities, are considered in the determination of household location, it is usually the distance from employment which often (as experience dictates it should) plays the largest role in making that determination. For many models, the location of employment is determined exogenously, i.e., it is "given" as an input to the particular model in question. Despite the extreme importance

of industrial allocation, modeling techniques for this sector are, for the most part, unsatisfactory. The significant influence of basic employment location and the present unreliable state-of-the-art, as compared to the extensive work in the area of residential location, necessitate the inclusion of a brief discussion on the problems of and possible approaches to the allocation of industrial employment.

VII. 5. A. The Problems of Industrial Employment Allocation

Effective industrial modeling is, at present, hindered by two major conceptual problems: (1) the location of a single industry vis-a-vis many industries; and, (2) inter-regional location vis-a-vis intra-regional location.

The basic allocation models forecast the spatial distribution of industrial employment within a region, given the projected growth for that region as a whole. Such models, trying to predict industrial location and movement at the macro level, have very little in common with those theories and models which restrict themselves to a single plant (industry). It is extremely difficult to transfer the principles derived from a study of the location process of an individual firm to the aggregated problem of the location of several types of industries. The difficulties arise on a number of fronts. First, the theory itself is by no means satisfactory. Too little is known, or at least quantifiable, about locational interdependence and the advantages of agglomeration. Relatively little is known

about the decision-making processes of firms, e.g., why a location is chosen which is not always the apparent optimum economic one. While such problems do exist, certain characteristics of the locational requirements of individual industries, such as transportation needs, water and waste facilities requirements, can be determined. However, the identification of the characteristics and the evaluation of alternative sites are obviously greatly compounded by any attempt to extend the analysis to several types of firms. A single industry or plant is concerned with only one set of locational requirements, whereas the peculiar needs of each of the industrial group must be considered when allocating several types of firms. In addition, a greater variety of sites is needed to meet the needs of several industries. Thus the difficult, though not impossible, task of locating a single industry provides some insight into the potential problems of a multi-industry modeling procedure.

The second major problem is that intra-regional industrial allocation differs markedly from the more usual one dealing with inter-regional location. The inapplicability of existing theory to the intra-regional case arises from the relative homogeneity within a region as contrasted with the heterogeneity among regions. Several examples will serve to demonstrate this point.

1. Transportation costs between the subareas of a region do not differ as greatly as between the regions themselves.

2. Wage rates tend to vary less within a region than among regions.

3. Interest rates differences are virtually nonexistent within any given region, while large discrepancies certainly do exist inter-regionally.

4. Because state and federal taxes are constant, there is little variation within any given region.

5. Institutional factors, such as union strength or business "climate", differ only slightly within a region, while among regions they much differ greatly, thus affecting the relative "attractiveness" among regions.

6. Utility rates (telephone, gas, electricity) vary little within a region, but here again, very greatly among regions and explain some of the attractiveness of some regions for certain types of industry.

Thus regions, from a locational viewpoint, tend to be quite homogeneous from within. Many of the regional differences or relative advantages normally incorporated in determining industrial location among several regions are inoperable from an intra-regional standpoint. Thus the number of variables which can be used for such an analysis are few and often difficult to quantify.

These two problems greatly reduce the utility or practicality of much of the industrial location theory and experience. Other problems such as difficulty in determining the industrial bid price for a particular parcel and the rather uncooperative nature towards

disclosing future expansion and location plans exhibited by most firms further contribute to the somewhat dismal state of the industrial allocation models.

VII. 5. B. Possible Courses of Action

The problems discussed in the preceeding section do not obviously circumvent any attempts at industrial employment allocation. Such allocation is crucial to all types of planning and development work and therefore will be done one way or another. While several such models are summarized later in this report, the present section is intended to discuss some of the more general aspects of such allocations.

The nature of the work being done and the degree of relative accuracy required by the Illinois Environmental Protection Agency mitigate, to a large degree, the problems encountered with basic employment allocation. There are several reasons which substantiate such a statement. First, the present level of sophistication at which air quality maintenance planning is being conducted does not warrant the determination of the exact site upon which a particular industry is expected to locate. For present planning purposes, from an air quality standpoint, identification of the general area in which location is likely to occur should suffice. At a much later time, when the analysis has advanced to a much more sophisticated state, more detailed models examining specific urban areas may be warranted.

Secondly, this "general area" criterion makes possible the use

of the more simplified procedures, such as the weighted overlay technique (discussed later in this report), for achieving relatively acceptable results. Such techniques are much easier to use and require considerably less expenditure of time and money than do the more complex models. Through local and county data the existing industries in the county can obviously be identified. Using the judgement and knowledge of local officials and planners, the relative growth or decline which these industries are experiencing can, to a degree, be also identified. Based upon land availability, capacity and density requirements, and zoning schemes, some general idea of which firms will be able to expand at their present sites and which will have to either relocate or add a small addition elsewhere, can be obtained. If relocation is necessary, it is highly unlikely that it would occur outside of the immediate vicinity. Since an extremely large percentage (90+%) of any county employment projection will be composed of new growth in existing industries, the allocation problem, from a simplistic point of view, is largely solved. It is the location of <u>new</u> industry which invokes many of the difficulties mentioned earlier in this section. The obvious problem is determining what amount of the projected growth is either entirely new industry (somewhat determined by the projection model used) to the region, <u>and also</u>, what amount of that growth will be comprised of new firms (subsidiaries or competitors) of the types presently existing in the region. The second category obviously may or may not locate near those already present.

The techniques discussed in this report reprecent operational procedures which are potentially capable of producing reasonable approximations of industrial employment location. For the present anticipated work of the IEPA these should provide acceptable results and a starting point from which to begin any additional investigation of this class of allocation models.

VII. 6. Model Summaries

VII. 6. A. Weighted Overlay Procedure

For lack of a better name this technique will be termed the weighted overlay procedure, although overlays in a strict McHargian sense are not employed. In his book, Design With Nature, Ian McHarg used transparent overlays with varying tonal hues and densities to assess the suitability of a particular land characteristic (e.g., slope, soil, surface geology) for development, preservation, etc. Although there are inherent problems with the McHarg procedure, such as the summation of absolute ordinal values, the technique is appealing and frequently applied.

While the procedure may theoretically be applied to any type of land use, the present discussion will be limited to the allocation of industrial activities. The first step involves the identification of these factors which are deemed, through local experience and location theory, to affect the location of manufacturing activities. A typical list may include the following factors:

(1) Highway accessibility

(2) Rail accessibility

(3) Airport accessibility

(4) Sewer and water facilities

(5) Slope

(6) Soil

(7) Availability of land for expansion

(8) Compatibility with surrounding land uses

(9) Business "climate"

(10) Tax rate

An additional site characteristic, which is mentioned in volume 4 of the EPA Guidelines for Air Quality Maintenance and Analysis, which should possibly receive consideration, is to view the site in the perspective as a limited, valuable resource which should be protected from misuse. Thus an additional factor to be considered is the potential harm to an area which may occur as a result of development.

The procedure may be described and summarized in the following steps.

(1) Estimate the total industrial activity to be located.

(2) Identify all potential sites for industrial development within a county. This would require a review of all pertinent local and county development plans and zoning schemes.

(3) Determine the relative importance (weight) of each factor with regard to industrial location.

(4) Evaluate each factor for each possible site and assign a

numerical value.

(5) Sum factor scores for each site and rank order the sites.

(6) Allocate industrial growth on the basis of site factor score divided by the summation of all site factor scores.

The procedure is most suited to allocation among sites within a particular metropolitan area.

This type of site characteristic evaluation is the basis of many of the more complicated allocation models such as the Hartford Area Traffic Study model. These models employ time-series data and regression analysis to determine the relative importance or weight of each factor considered. However, the above described form is operationally more appealing from a time and cost standpoint. Very little expenditure of either is required, and, in view of the accuracy problems of the regression models, the resulting allocations would be expected to be satisfactory for any type of gross land use allocation study undertaken by the Environmental Protection Agency.

The procedure possesses further appeal in the sense that it considers, even demands, input from local-county officials and planners. For example, available tracts which are known to be marked for future development would obviously be excluded from consideration. Other factors not specifically considered, such as special local concessions or agreements, could be incorporated into the allocation process. Furthermore, local input can be incorporated in the factor, "compatibility with surrounding land uses." Used as a control factor, a zero score (indicating incompatibility)

would circumvent any further evaluation and consideration of a given site.

Obviously a certain amount of care and good judgement must be exercised in making allocations to those sites with relatively high scores. If the allocation study is done on a county basis, it would not be feasible to evaluate each site within each urban and rural area separately. Rather, several areas within and around one urban area may be classified as a site, and the aggregate evaluated. If this site received the highest score, one would not necessarily allocate the total amount of industrial growth required to "fill" the entire site. What this points to is the fact that the size of the site, and the amount of land within that site which is likely to be developed, must be considered. This involves a review of the locations of existing industries, their likelihood for expansion within an area, and the holding capacities or densities of the sites. Despite such caveats, the weighted overlay procedure represents an extremely operational approach to allocation land uses at level of accuracy suitable for many investigative studies.

VII, 6. B. The Chicago Area Transportation Study Model (CATS 1965)

This is not a mathematical model in the sense that other models may be so described. The industrial location involves a good deal of judgement by the study team. The areal units of interest are the region, the ring, district and traffic zone in order of decreasing size. For each land use zone the future inventory of land uses is

extrapolated from the initial inventory according to rules, modified by judgement, specific to the kind of use. Six types of land uses are recognized in the model: residential, commercial, manufacturing, transportation, public buildings, public open space and streets. Vacant land is classified as residential, commercial or insustrial, according to local zoning ordinances. Unusable land is also inventoried and deleted from consideration.

The initial land use pattern of each land use zone is modified according to the following steps:

(1) Specific parcels of land in certain zones are designated for conversion to public open spaces and transportation, based primarily upon existing plans of public agencies for such development.

(2) Based upon announced private plans and staff judgement, commercially zoned vacant land in certain districts is designated for shopping centers and heavy commercial uses.

(3) Residentially zoned vacant land is designated for residential use, the amount dependent upon the location of the zone and its residential holding capacity at existing or modified densities. The percentage of a zone's holding capacity to be filled by a target year (1980) is a function of distance from the Central Business District, with modifications based on staff judgements.

(4) Space for streets, local commercial facilities, public buildings, and recreation is set aside in each residentially oriented zone based upon application at per capita norms to the estimated 1980 population of each zone as determined in step (3).

(5) Industrially zoned vacant land is designated for manufacturing use, based upon the location of the zone and its manufacturing holding capacity. Trends in net employment density over time and by distance from the CBD, along with the amount of industrially zoned space determine a zone's holding capacity. The percentage of this capacity to be filled by 1980 is a function of distance from CBD, with modifications based on staff judgements.

(6) When net activity density and acreage have been explicitly determined for each zone, the implied population and employment totals for each zone can be calculated. These are summarized for the study area as a whole and compared to independent projections of the area's population and employment. The land use forecast of acreage occupied is then systematically modified so as to reconcile the implied activity totals with the independent projections.

Thus the CATS model is seen to be largely based upon local knowledge and input, and the judgements of the staff members. The subjective nature of the model had been the source of varied critics of the procedure. However, it does represent a systematic, relatively inexpensive approach to the problem of allocating the projected population and employment totals of an area. The potential inaccuracies of such an approach must be weighed against the relative ease of use and savings of time and money associated with its application.

VII. 6. C. Residential Growth Model - South Bend Urban Area Transportation Study (Bartholomew and Assoc., 1970)

VII. 6. C. a. Introduction

The model generated for the South Bend area represents a compromise between the "subjective" model which distributes residential growth on the basis of professional judgement, and the more "sophisticated" approaches which employ rather involved procedures. The model distributes residential growth to various areal units based on their accessibility to employment concentrations as well as the amount of residential land available and the residential density in each areal unit. The model does provide for the inclusion of input from professional planners and local officials. Future population growth is distributed on an incremental basis whereby the output of each incremental distribution provides input to the next distribution.

VII. 6. C. b. Mathematical Formulation of the Model

The distribution process involves the allocation of the forecasted population growth to areal units (zones) proportional to zonal distribution ratios, DR_i. The ratio is a function of the relative zonal accessibility to employment opportunities, the availability of residential land for development, and population densities and capacities. Accessibility is defined as the potential for opportunities to interact. Differing from some conventional definitions, accessibility is a measure of the intensity of interaction, rather than simply a measure of the ease of interaction.

Mathematically stated, accessibility at Zone 1 to an employment activity in Zone 2 is directly proportional to the size of the activity and inversely proportional to the travel time between

Zones 1 and 2. The total accessibility of employment is Zone 1 is a summation of the accessibilities to each of the individual zones within the study area.

A modified version of the model is currently being employed in a transportation study for the Evansville, Indiana metropolitan area. For the original model a regression equation relating spatial separation between two zones and travel time between the zones was generated. This involved a regression technique which considered various components of "friction" to travel other than simply physical separation. Although a significant relationship was established (R = 0.99), it is not certain whether this technique can provide accurate future projections. For this reason the modified approach is recommended and discussed below. The mathematical formulation of accessibility is:

$$ACC_i = (E_j/F_{ij}) \qquad (1)$$

where: ACC_i = accessibility of Zone i

E_j = employment at Zone j

F_{ij} = travel time between Zones i and j.

Distribution ratios (DR_i) represent relative measures of zonal population potential, and are a function of the absolute population potential of various zones. The latter are referred to as distribution factors and are difined as:

$$DF_i = ACC_i \times PC_i \qquad (2)$$

where: DF_i = distribution factor for Zone i

ACC_i = accessibility of Zone i

PC_i = population capacity of Zone i.

Population Capacities are found by solving the following equation:

$$PC_i = LA_i \times PR_i \times R_i \times D_i \qquad (3)$$

where: LA_i = total undeveloped land in Zone i

PR_i = percent of total land in Zone i that is allocated to residential development

R_i = percent of total land in Zone i that is expected to be residentially developed

D_i = population density in Zone i.

The distribution ratios are obtained by dividing the distribution factor for each zone by the summation of the distribution factors for all zones, as shown below:

$$DR_i = DF_i / \sum_j DF_j \qquad (4)$$

where: DR_i = distribution ratio for Zone i

DF_j = distribution factor for Zone j.

Population is then allocated to various zones by multiplying the zonal distribution ratios by the total aggregate population growth according to:

$$Z_i = TPG \times DR_i \qquad (5)$$

where: Z_i = population growth allocated to Zone i

TPG = total aggregate population growth

DR_i = distribution ratio for Zone i.

VII. 6. C. c. Data Requirements

The input data to the residential growth model are divided into two categories. The first represents data directly obtainable from planning files, past surveys, etc.

1. Total acres - total zonal area in acres

2. Residential acres - total zonal residential area in acres

3. Agricultural, wooded and vacant - area by zone in acres

4. Zonal population

5. Zonal employment

6. Interzonal travel times - in minutes

The second category consists of forecasted data:

1. Population Density - the zonal net population expressed in persons per residential acre

2. Percent Allocation - the percentage of agricultural, wooded and vacant land allocated to residential purposes. This variable would consider factors which would decrease the total amount of land available for residential development such as streets, sidewalks, utility right-of-ways, etc.

3. Percent Expected - this variable represents the embodiment of local knowledge and expertise into the allocation process. Included would be analysis of past trends, known or planned developments which will occur, etc. By employing specific knowledge of conditions in the area, more accurate allocations of residential growth should be generated. This variable represents an exogenous control measure restraining the extent of new residential development in each zone and therefore becomes a measure of control over the mathematical procedure which does not recognize such factors as amenities, tastes, personal "inertia", etc.

4. Zonal Employment - the estimate of the future employment in

each zone. This obviously requires the application of some employment allocation procedure to the estimate of total employment growth during each year of the residential growth forecast procedure.

VII. 6. C. d. The Allocation Process

The incremental (year-to-year) allocations may be performed in three stages to evaluate various policy alternatives:

1. Unrestrained - no consideration is given to limiting the population distribution to zonal physical capacity set by the judgement decisions on such matters as density and acreage of land available for residential development.

2. Capacity restrained - consideration is given to the total zonal capacity for residential development but the estimate of the percent of the total available residential land to actually be developed within the first incremental period is not recognized.

3. Adjusted capacity restrained - the estimate of the amount of land that is expected to be developed is incorporated as a restraint. The Stage 1 and 2 distributions are not essential to the model's operation. They merely provide some interesting insight into the effects of alternative housing policies. The unrestrained distribution provides an indication of where people would desire to live if sufficient housing and all other associated facilities could be provided within the zones. Thus a zone with an unrestrained distribution of residential growth greatly in excess of the zonal capacity would suggest that a higher development density might be appropriate and should be given consideration for the next

incremental forecast.

For the restrained distribution, it is necessary to check for zones where zonal capacity is less than the zonal population allocations. Where there is an excess of zonal population allocation, the excess is redistributed to other zones possessing an excess of capacity, using the zonal distribution ratios described earlier. The resulting population allocations are again compared to the zonal capacities, and the process repeated until the zonal allocations are less than or equal to the corresponding zonal capacities.

VII. 6. C. e. Conclusion

The South Bend model represents a simplified, operational approach to residential growth distribution. The relatively simple computations and the types and availability of data needed as input, suggest that the distribution could be accomplished with a relatively small expenditure. The importance of this aspect can only be appreciated after one examines the enormous amounts of time and money which must be spent in acquiring data for the more complicated models, some of which are mentioned later in this report.

As mentioned earlier the residential distribution model is dependent upon a projection and distribution of future employment. From this distribution, residential growth may be allocated on the basis of accessibility to employment activities.

The model provides several means for input and modification by planners, public decision-makers, local civic leaders, etc. The

most evident perhaps is in the variables in equation (3). Obtaining values for the percent of land that is expected to be residentially devoloped most certainly calls for input from a number of sources. The analyst would obviously want to consult city, county and regional planning commissions as to what types and amounts of development have occurred, and are expected to occur in the future. Manipulation of the zonal population density will enable the effects of varying densities upon residential distribution to be examined. Thus, various local and regional zoning schemes may be substituted into the model for determining an optimal allocation pattern.

The model is a flexible method which may be applied on either the macro or micro scale. On the macro scale, projected county residential growth could be allocated to various zones, with the zones being either entire cities or large areas of rural land. If all that is desired is knowledge of the general area where residential growth would occur, then this type of scale would be acceptable. This would obviously require that employment be distributed either on this type of scale or aggregated into figures representing several sites within one of the residential zones.

If, on the other hand, more precise locational information is desired the model could be applied on the micro scale, as is being done in South Bend and Evansville. In this case, the zones would be areas within a city or the surrounding rural areas. This type of allocation would provide more detailed estimates of where residential

growth is expected to occur. This type of locational information would be needed for any transportation study attempting to estimate the number and location of trips generated by economic growth. The most satisfactory approach would perhaps be a two stage allocation, first on the macro scale, then followed by a micro distribution. This would then provide information for use at both the regional and local levels.

VII. 6. D. The Lowry Model (Lowry 1964)

Developed by Ira Lowry in 1962-63 as a part of a modeling system for the Pittsburgh Comprehensive Renewal Program, the Lowry Model makes two basic assumptions about the factors governing the location of activities in regions and urban areas. First, it is assumed that an individual's choice of residential location is strongly influenced by the location of his place of work. Secondly, it is assumed that economic activities can be divided into two categories on the basis of their locational requirements.

The urban spatial structure incorporated into the model is composed of three groups of activities:

1) A basic sector

2) A retail sector

3) A household sector

The basic sector includes industrial, business and other activities whose locations are assumed to be unconstrained by local markets and population distributions. The employment levels and locations of the basic sectors are assumed as "given", and therefore, determined

outside the model. Included in this category are those activities producing intermediate products for industries serving local markets.

The retail sector includes all those activities directly dependent upon local resident population and purchasing power. The employment levels and locations in this sector are considered endogenous whose values are determined within the model. This activity sector is broken down into neighborhood, local, and metropolitan clusters.

VII. 6. D. a. Formulation of the Model

The following notation is that used in the original model:

A = Area of land

E = Employment

N = Population (number of households)

T = Index of trip distribution

Z = Constraints

U = Unusable (land)

B = Basic sector

R = Retail sector

H = Household sector

k = class of establishment within the retail sector

m = number of classes of retail establishments

i,j = subareas, called tracts

n = number of tracts (i,j=1,2,...,n)

The structural parameters are indicated by:

a^k = retail employment coefficient related to class k or population serving ratio

b^k = retail employment scale factor related to class k

c^k, d^k = shopping trip weight factors related to class k ($c^k + d^k = 1$)

e^k = retail employment density ratio related to class k

f = the reciprocal of the labor force participation rate

g = population scale factor.

The following data are given as input to the model:

1) The level, categorization and allocation of basic employment

2) An inventory of unusable, basic, retail and residential land

3) The interzonal network - used to calculate distances and trip distribution indices

4) Structural parameters

5) Constraints on densities of residential development, on the minimum site of a cluster of retail activities, and on the amount of land available in each tract. Tracts are defined to be 1-mile square.

The allocation procedure is done in a series of iterative steps. First, using the level and location of basic activities as given, the model derives the related basic employee population which is then allocated spatially over the study area. It then derives the retail activities required to serve this amount of basic population and allocates this employment to service areas. From this level and location of service employment, the model derives the dependent population related to this new employment force. This additional population is then allocated, which in turn demands new services,

which in turn create more jobs, and so on. The model iterates in this way until a stable solution is achieved for a given input of basic employees.

VII. 6. D. b. Mathematical Formulation

(A) Given the basic employment, total population is predicted by:

$$N^* = f\, E^B (1 - f \sum_{k=1}^{m} a^k)^{-1} \qquad (1)$$

For the first iteration $E_j = E_j^B$; thus total population is:

$$N = f \sum_{j=1}^{n} E_j^B$$

(B) Total employment is calculated by:

$$E_j = E_j^B + \sum_{k=1}^{m} E_j^k \qquad (2)$$

(C) Residential land is next found by:

$$A_j^H = A_j - A_j^U - A_j^B - A_j^R$$

For the first iteration, $A_j^B = 0$ because no retail activities have yet been allocated.

(D) Population is (re) allocated by the following potential submodel:

$$N_j = \frac{N \sum_{i=1}^{n} \frac{E_i}{T_{ij}}}{\sum_{j=1}^{n} \sum_{i=1}^{n} \frac{E_i}{T_{ij}}} \qquad (3)$$

(E) Population for each tract N_j is tested against a maximum density constraint Z_j^H by:

$$N_j = Z_j^H A_j^H \qquad (4)$$

if the constraint is violated, the surplus population is reallocated until the constraint is satisfied.

(F) Total retail employment by class is derived by the following equations:

$$N = \sum_{j=1}^{n} N_j \qquad (5)$$

$$E^k = a^k N \qquad (6)$$

(G) Retail employment by each category k is (re) allocated by the following potential submodel:

$$E_j^k = \frac{E^k \sum_{i=1}^{n} \left(\frac{C^k N^i}{T_{ij}}\right) + d^k E_j}{\sum_{j=1}^{n} \left(\sum_{i=1}^{n} \left(\frac{C^k N^i}{T_{ij}^k}\right) + d^k E_j \right)} \qquad (7)$$

(H) The minimum size constraint is now tested to ensure that retail employment in any tract E_j^k is greater than some minimum size limit Z^k

$$E_j^k \geq Z^k \qquad (8)$$

if the constraint is not satisfied, the retail employment of tract j is set to zero and the value reallocated.

(I) Retail land needed for each tract is next calculated by:

$$A_j^R = \sum_{k=1}^{m} e^k E_j^k \qquad (9)$$

(J) The land for retail activities is tested against the amount actually available

$$A_j^R \leq A_j - A_j^U - A_j^B \qquad (10)$$

if $A_j^R > A_j - A_j^U - A_j^B$, A_j^R is set equal to $A_j - A_j^U - A_j^B$, thus allowing for over crowding, and at the same time, setting $A_j^H = 0$. Thus retail

activities have priority over residential uses. If households have been allocated to such a tract, they will be removed by the residential density constraint of the next iteration.

(K) Total population N used by the model as an independent variable is compared with the total population N*, predicted by the model in equation (1); N is found by:

$$E_j = E_j^B + \sum_{k=1}^{m} E_j^k \qquad (2)$$

$$N = f \sum_{j=1}^{n} E_j \qquad (11)$$

If the two values are different, the model is run in an iterative fashion until they converge on a stable value. In the Pittsburgh study, only three iterations were necessary.

VII. 6. D. c. Conclusion

The Lowry model is perhaps the most copied or modified allocation model presently in existence. Its basic features have been incorporated, for example, in the Bay Area Simulation Study Models (BASS I, BASS III) and The Projective Land Use Model (PLUM). These and other modified models are all disproportionately more costly than the accuracy improvements which they claim to possess. The Lowry model concept has, in addition, been applied a number of times in Great Britain.

The model is extremely well documented and is one of the simplest models to construct in terms of data requirements. Lowry has done extensive research on determining values for the trip distribution indexes (T_{ij}) and the shopping trip weight factors

(c^k and d^k).

The number of tracts analyzed in the model is largely constrained by the computer capacities to which the program is to be fitted, data availability, and the importance of locational exactness for which the output is to be used. The model, however, is entirely feasible for use at the county or regional level. Once the basic employment has been exogenously determined, the allocation task would simply involve n-separate allocation problems in each of the basic employment areas (cities, counties, etc.)

VII. 6. E. The Empiric Model (Traffic Research Corporation 1967) (Hill 1965)

The function of the model is to distribute or allocate externally-supplied growth forecasts of the various activities among the sub-regions (groups of counties) in the Boston Metropolitan Region in steps of 5 or 10 year periods. The model is based upon the concept that the development patterns of urban activities are interrelated in a systematic manner which provides a reasonable basis for their prediction.

The model predicts the amounts of several urban activities (located variables) in each subregion at the end of a given forecast period. Those located variables are:

(1) white collar population

(2) blue collar population

(3) retail and wholesale employment

(4) manufacturing employment

(5) all other employment

The amount of each located variable is related to the presence or absense of certain locator variables (e.g., accessibility, sewer facilities, etc.) The change in the sub-regional share of a located variable is proportional to: (A) The change in the sub-regional share of all other located variables in the sub-region; (B) The change in the sub-regional share of a number of locator variables in the sub-region; and, (C) The value of the sub-regional shares of other locator variables. This concept is expressed as:

$$\Delta R_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} \Delta R_j + \sum_{k=1}^{m} b_{ik} (Z_k \text{ or } \Delta Z_k)$$

where: i or j = 1,2,...,i, j,...,N: number of the located variable (a total of N equations)

k = 1,2,...,k,...,M: number of the locator variable

ΔR_i or ΔR_j = change in the level of the ith or jth located variable over the calibration or forecast time interval

Z_k = level of locator variable k at the beginning of the calibration or forecast time interval

ΔZ_k = change in the level of the kth locator variable over the calibration or forecast time interval

a_{ij}, b_{ik} = coefficients expressing the interrelationships among variables.

There is one such equation for each located variable. The a and b coefficients are determined by simultaneous regression analysis of data from two past points in time. Using the values obtained for a and b and projected exogenous values of Z_k or ΔZ_k, the system

of equations of the form of equation (1) are solved simultaneously to estimate future change in the level of each located variable for each sub-region. These changes in share are then added into the shares held by each sub-region at the beginning of the forecast period to obtain a total relative value of the sub-regional share at the end of the forecast period. To obtain absolute values these relative values are then multiplied by the externally supplied forecast control figure.

VII. 6. E. a. Calibration and Validation of the Model

Calibration and testing of the model revealed that it did indeed explain a high percentage of the change throughout the 1950-60 decade in the Greater Boston Region.

VII. 6. E. b. Conclusion

The development, calibration and test work have demonstrated the model to be both practical and operational. The test model has achieved a level of accuracy that compares favorably with the accuracy obtainable from the best of present-day forecasting models. Continuing research is being done in hopes of developing a model which permits forecasts of up to 10 activities for up to 600 sub-regions. While the model does require detailed data on the locator variables, it still represents one of the more accurate and operational models presently in existence.

VII. 6. F. Comprehensive Allocator of Systems, Services and Regional Activities (CASSANDRA) (Jhaveri, 1973)

The CASSANDRA model, developed by the Northeastern Illinois

Planning Commission staff, allocates the population and employment activities of the Northeastern Illinois Region to its constituent townships for the purpose of determining the regional distribution of activities and demand for land for the years 1980 through 2020. Rather than predicting this future distribution under the assumption of little or no public intervention, as most models do, the proposed allocation model emphasized the incorporation of goals of urban form and public policies and objectives as adopted by NIPC. For example, the Regional Comprehensive General Plan places significant emphasis on mass transit and high density of growth on the major transportation corridors, and must be considered as an input to the model.

VII. 6. F. a. Activities to be Forecasted by the Model

The model generates land use forecasts of the following activities:

1) High income households
2) Middle income households
3) Moderate income households
4) Low income households
5) Manufacturing employment
6) Transportation, communication, utilities
7) Wholesale trade
8) Retail trade
9) Finance, insurance, and real estate
10) Local government employment
11) Institutions

12) Federal and state employment

VII. 6. F. b. Model Structure and Calibration

Using data from two points in time (e.g., 1960 and 1970), the total activity (e.g., number of households) (H) in any sub-area i at time t-1 can be classified into stable (H^s) and relocating or moveable activity (H^m).

$$H_i(t-1) = H_i^s(t-1) + H_i^m(t-1) \qquad (1)$$

In addition to the activity expected during the calibration period, t-1 and t, the increase or decrease in activity for the region as a whole must be considered. Total activity to be relocated during the calibration period for the region $H^r_{(t)}$, is the sum of moveable and net growth activity in the region.

$$H^r_{REGION\ (t)} = \sum_i H_i^m\ (t-1) + \Delta H_{REGION} \qquad (2)$$

The relocating activity in the region as initially assigned to sub-areas in proportion to their share of certain variables Z^k, which are assumed to affect an activity's location (e.g., open space, employment distribution).

$$H_i^k\ (t) = (Z_i^k\ (t)/\sum_i Z_i^k)\ H^r_{REGION\ (t)} \qquad (3)$$

The relocating activity $H_i^k\ (t)$ initially allocated to sub-area i searches for locational opportunities O_j in sub-area j, including those within itself. Locational opportunities are comprised of the level of activity relocated from $j(H_j^r)$, plus the potential opportunities suggested by vacant land and structures, (O_j^p).

$$O_j\ (t) = H_j^r\ (t) + O_j^p\ (t) \qquad (4)$$

Using an intervening opportunity model, such as the one used extensively in the Chicago Area Transportation Study, the probability of relocation from areas of initial allocation (i) to where locational opportunities are available (j) is determined, P_{ij}^{k}.

$$P_{ij}^{k} = (1-L)^{o} - (1-L)^{o+oj} \qquad (5)$$

where: o = the sum of the locational opportunities between i, and areas closer to i than is j.

L = the influence of a variable on the distribution of activities among opportunities surrounding the variable.

Theoretically, an L value is the probability that an activity starting from the concentration of a given variable and reaching for locational opportunities, would find any given opportunity acceptable. Such values are typically generated in origin-destination surveys.

The level of activity relocated to sub-area j from the zone of initial allocation by variable Z^{k} is:

$$H_{ij}^{k}(t) = P_{ij}^{k}(t) \times H_{i}^{k} \qquad (6)$$

Total activity received by sub-area j during the calibration period, if variable Z^{k} were the only consideration, is the sum of activity received by j from all sub-areas, including j.

$$H_{j}^{k}(t) = \sum_{i} H_{ij}^{k}(t) = \sum_{i} P_{ij}^{k}(t) \times H_{i}^{k}(t) \qquad (7)$$

H^{k} is a vector of size j = 1,2,...,m sub-areas and represents the predicted distribution of the activity based on Z^{k} variables, considered one at a time.

Because the actual distribution of the relocating activity H

is known, it may be regressed against the predicted distribution of the activity by the Z^k variables H_j^k. For the purpose of regression analysis let the distribution of H_j^* be defined as the vector Y and the distribution of H_j^k be defined as the vectors $x_1, x_2, \ldots, x_n$.

$$Y = a + b_1X_1 + b_2X_2 \ldots b_nX_n \qquad (8)$$

Multiple linear regression gives the values of the coefficients, which are the relative weights of the distributions in explaining, the actual distribution of the relocating activity during the calibration period.

VII. 6. F. c. Forecast

The activity forecasting process is essentially the same as the calibration except that the coefficients associated with the independent vectors $b_1, b_2, \ldots, b_n$ are given from the preceeding period.

Once the independent vectors $x_1, x_2, \ldots, x_n$ are estimated for time t+1, the estimated level of the relocated activity in any sub-area is the sum of the $H_j^k(t+1)$ vectors multiplied by the corresponding coefficients.

The estimated level of total activity in any sub-area j in t+1 is the sum of the relocated activity H received plus the stable activity from time t.

VII. 6. F. d. Conclusion

The CASSANDRA model can be termed operational in the sense that it was designed with the idea of the using existing natural

resource and land use data files of the Commission, the census tapes, and employment statistics from state and federal sources. Because the model is presently being utilized by a regional planning agency in the State of Illinois, there is a certain appeal to incorporating the method on a state-wide basis. Those who are responsible for the development and modification of the model would hopefully be readily available for consultation and local technical assistance.

Because the results of the regression estimation procedure were not reported, no evaluation of the model's relative accuracy can be made in this review. Similarly, suitability of the approach to other areas would be dependent upon the ability of the regression equations to at least replicate past conditions. Thus it would behoove one to examine the statistical significance of the regression coefficients. If tests confirm that the regressions are significant, and appear reasonable, the coefficients may be used for allocating future activity in the region if we assume that the distribution behavior of an activity in the future assigns the same weight to the various variables as was observed in the past.

VII. 6. G. Developing and Testing an Intra-Regional Model (Putnam 1970)

Undertaken as a part of the U.S. Department of Transportation's study of the Northeast Corridor of the United States, this study initially incorporated an interrelated set of three models: an econometric model used as input to an input-output model, which was

then used as input to an intra-regional location model.

Using the concepts of economic base theory, the model discriminates between two classes of employment. Those linked primarily to local population-serving activities (services and retail) are considered non-basic. Those types of employment which primarily serve other local and non-local industries are considered as basic. The procedure first makes separate projections of basic, non-basic employment and population, and, using these totals, projects total land use for counties within the study area.

VII. 6. G. a. The Basic-Employment Equations

The four types of basic employment considered are (1) Agriculture, Forestry, Fisheries; (2) Mining; (3) Manufacturing; and, (4) Transportation, Communication and Public Utilities. Basic employment is postulated to be a function of (A) the spatial distribution of suppliers of inputs; (B) the spatial distribution of consumers; (C) transport facilities; (D) basic employment in the previous time period; (E) total employment in the previous time period; and, (F) gross population density in the prior time period.

Two accessibility measures, one to imputs and one to market are used in making employment projections. Accessibllity to suppliers is of the form:

$$A^{s}_{i,g,t-1} = \sum_{j} \left(\frac{\sum_{h} (a_{gh} P_{h} E_{j,h,t-1})}{Z^{f}_{i,j,t-1}} \right) \qquad (1)$$

where: $A^{s}_{i,g,t-1}$ = a measure of the accessibility at time t-1 of area i to inputs required by industry group g

$E_{j,h,t-1}$ = employment in group h in area j at time t-1

P_h = productivity (output/employee) for industry group h

a_{hg} = input-output coefficients of inputs from sector h per unit of sector g output. (O.B.E.)

$Z^f_{i,j,t-1}$ = freight "impedance" between areas i and j at time t-1. Includes costs of shipment, time, modal split.

Accessibility to market access is of a similar form:

$$A^m_{i,g,t-1} = \sum_j \left(\frac{\sum_h (a_{gh} P_h E_{j,h,t-1})}{Z^f_{i,j,t-1}} \right) \qquad (2)$$

Equation (2) differs from (1) by use of a_{gh} which represents the number of units of output from sector g used as input by sector h to produce one unit of sector h output. The numerator of (2) thus represents a meauure of potential demand for the outputs of sector g.

The various variables are combined in a linear function as below:

$$E_{i,g,t} = b_1 A^s_{i,g,t-1} + b_2 A^m_{i,g,t-1} + b_3 E_{i,g,t-1} + b_4 E^{T'}_{i,t-1} + b_5 (P_{i,t-1}/L_i) + b_6 \qquad (3)$$

where: $E_{i,g,t}$ = employment of insustrial sector g in area i at time t

$A^s_{i,g,t-1}$ = accessibility at time t-1 of area i to inputs required by sector g

$A^m_{i,g,t-1}$ = accessibility at time t-1 of area i to markets which consume the outputs of sector g

$E^{T'}_{i,t-1}$ = total employment minus sector g employment at time t-1 in area i

$P_{i,t-1}$ = population residing in area i at time t-1

L_i = land area of area i.

Using a single-equation least squares regression procedure, the following coefficients of determination were obtained for the study region:

Basic Sector 1 - Agriculture, Forestry, Fishing	$R^2 = 0.92$
Basic Sector 2 - Mining	$R^2 = 0.74$
Basic Sector 3 - Manufacturing	$R^2 = 0.99$
Basic Sector 4 - Transportation, Communication, Public Utilities	$R^2 = 0.97$

VII. 6. G. b. The Non-Basic Employment Equation

Four types of employment comprise the non-basic class: (1) Retail trade; (2) Finance, Insurance, Real Estate; (3) Services; and, (4) Government. Non-basic employment is postulated to be a function of (A) total population in an area, which is essentially a measure of the market for the output of the employment; (B) total employment in an area, which is also a market measure; and, (C) the amount of that particular type of employment in the area in the prior time period.

Non-basic employment of a given sector is determined by the following linear function:

$$E_{i,g,t} = c_1 E_{i,g,t-1} - c_2 P_{i,t} + c_3 E^t_{i,t} + c_4 \quad (4)$$

where: $E_{i,g,t-1}$ = employment of non-basic sector g in area i at time t-1

$P_{i,t}$ = population of area i at time t

$E^t_{i,t}$ = total population in area i at time t.

The results of estimating the equations were as follows:

Non-basic Sector 1 - Retail trade $R^2 = 0.995$

Non-basic Sector 2 - Finance, Insurance, Real Estate $R^2 = 0.998$

Non-basic Sector 3 - Services $R^2 = 0.99$

Non-basic Sector 4 - Government $R^2 = 0.91$

Commercial (wholesale trade) employment was determined as being neither wholly basic nor wholly non-basic. This employment level is postulated to be determined by the same factors as those which affect the wholly non-basic employment plus simultaneous accessibility to total employment (market and suppliers). The estimating equation is:

$$E_{i,5,t} = d_1 A^t_{i,t} + d_2 E_{i,5,t-1} = d_3 P_{i,t} + d_4 E^t_{i,t} + d_5 \qquad (5)$$

where: $A^t_{i,t-1}$ = total accessibility to employment of area i at time t-1

$$\sum_j \left(\frac{\sum_g P_g E_{j,g,t-1}}{Z^f_{i,j,t-1}} \right)$$

R^2 for the above estimation was 0.97.

VII. 6. G. c. The Population Estimates

Population in an area at time t is postulated to be a function of: (1) the population of an area at time t-1; (2) the number of births in an area since time t-1; (3) the number of deaths in an area since time t-1; and, (4) the net migration into an area since time t-1. The number of births and deaths are related to per capita personal income in an area, gross population density in an area, and prior births or deaths in the area. Regression techniques were

employed in a manner similar to that used in forecasting employment.

VII. 6. G. d. The Land Use Forecasts

Forecasts of three categories of land use were made: (1) industrial; (2) commercial; and, (3) residential. Accessibility, used as being reflective of land desirability, and density of present use are employed as the independent variables.

The industrial land use forecasting equation is:

$$U_{i,I,t} = e_1 E_{i,3,t} + e_2 (E^t_{i,t}/L_i) + e_3 A^t_{i,t} + e_4 \qquad (6)$$

The equation for forecasting commercial land use is:

$$U_{i,c,t} = f_1 (1.0/L_i) + f_2 (E_{i,5,t})^{f_3} + f_4 \qquad (7)$$

The equation for forecasting residential land use is:

$$U_{i,R,t} = k_1 (P_{i,t}/L_i) + k_2 \pi_i + k_3 A^p_{i,t} + k_4 \qquad (8)$$

where: $U_{i,I,t}$ = percent of industrial land in area i at time t

$U_{i,c,t}$ = percent of commercial land in area i at time t

$U_{i,R,t}$ = percent of industrial land in area i at time t

π_i = 0, when $(P_{i,t}/L_i)$ is less than 10,000

= 1, when $(P_{i,t}/L_i)$ is greater than 10,000

$E^t_{i,t}$ = total employment in area i at time t

L_i = land area of area i

$A^p_{i,t}$ = total accessibility to residential population of area i at time t

$$\sum_j (P_{j,t}/Z^p_{i,j,t})$$

$A^t_{i,t}$ = total accessibility to employment of area i at time t

$P_{i,t}$ = population of area i at time t.

The results of estimating these equations were:

Industrial land use by county	$R^2 = 0.75$
Commercial land use by county	$R^2 = 0.70$
Residential land use by county	$R^2 = 0.73$

VII. 6. G. e. Conclusion

While there are definite areas in which the model can and should be further developed, it does appear to produce reasonable accurate and reliable forecasts. Comparison of actual and predicted values of total employment and population, to a large degree, substantiated the reasonableness of the model forecasts.

The computer program of the model has been made operational and easy to use at the University of Pennsylvania's Computer Center. Obviously, application of the model to any study region would require calibration and testing, with no inherent guarantee of the reproducibility of the high R^2 values obtained in the present study.

The largest obstacle to incorporation of the model is the requirement of determining the accessibility and travel costs values of a particular region. However in regions in which any type of detailed transportation planning work which utilizes such measures has been conducted, this represents only a minor obstacle. The remaining data requirements appear to be easily within the reach of any public agency concerned with population and employment growth. In light of the above comments, it is felt that the model most certainly represents an operational procedure for making reasonable accurate estimates of future population and employment

and their associated land use allocations.

VII, 6. H. A Model for Allocating Economic Activities into Sub-Areas in a State (Voorhees and Assoc., 1966)

Prepared for the Connecticut Interregional Planning Program, the model allocates state aggregate totals of economic activities to 169 towns (sub-areas) within the state.

In order to estimate sub-area economic growth within a state it is necessary to:

(1) Relate the sub-area's development to development in the parent region as a whole;

(2) "Weight" its growth in relative terms, i.e., in terms of departure from the parent region norm;

(3) Examine the characteristics of its growth patterns; and

(4) Evaluate its changing position with regard to its ability to hold and attract persons and industries.

The allocation model is based upon the shift analysis framework. (For a detailed discussion of this technique see Perloff, 1961).

VII, 6. H. a. Model Formulation

The following notation will be used in the model's formulation:

E = level of activity (employment, population)

i = sector of activity

j,v = geographical unit (town)

o,t,t+1 = time periods.

The net change in population or employment in a sub-area between two points in time is:

$$N.C. = E_{ijt} - E_{ijo} \qquad (1)$$

This net change can be viewed as being composed of two components. This first is the net change which occurs if the sub-area change in any sector equals the rate of change if that sector in the state as a whole. The Proportional Share (P.S.) is given by:

$$P.S. = E_{ijo}\left(\frac{E_{it} - E_{io}}{E_{io}}\right) \qquad (2)$$

The second component of change is the amount due to the growth rate of the jth sub-area in the ith sector being faster or slower than the growth rate in that sector in the state as a whole. This differential or competitive shift (D.S.) is written as:

$$D_{ij} = E_{ijo}\left(\frac{E_{ijt} - E_{ijo}}{E_{ijo}} - \frac{E_{it} - E_{io}}{E_{io}}\right)$$

$$= E_{ijt} - E_{ijo}\,(E_{it}/E_{io}) \qquad (3)$$

From (1), (2), and (3) it can be shown that:

$$P.S. + D.S. = N.C.$$

Because the differential shift relates each sub-area to the state's performance, (faster growing towns will have positive values of D_{ij}, slower growing towns will have negative values of D_{ij}), it can be seen that:

$$\sum_{j} D_{ij} = 0$$

VII. 6. H. b. The Activities

Six employment sectors and three sub-groups of population (by income) form the nine activities with which the model is concerned.

The employment sectors used are:

(1)	Manufacturing	(SIC 19 - 39)
(2)	Retail and wholesale	(SIC 50 - 59)
(3)	Personal services	(SIC 63 - 88)
(4)	Business and professional services	(SIC 60 - 62, 89)
(5)	Construction	(SIC 15 - 17)
(6)	Other	(SIC - all other)

VII. 6. H. c. The Land Use Model

Using data on the nine sectors for two points in time, the differential shifts are calculated. The land use model predicts the differential shift of each sector in each sub-area in terms of both endogenous and exogenous variables. The differential shift of one sector (endogenous) is predicted in terms of differential shifts (endogenous) of other related sectors. The model is formulated as a set of simultaneous equations where each equation describes one sector. This generalization of the multiple regression technique allows for the incorporation of several dependent variables into the system of equations, thus allowing the dependent values to interact much in the same way they do in actual conditions. Thus the interdependencies in the location of various economic activities can be treated explicitly in the simultaneous equations approach. In this system of equations, a variable which is dependent in one equation becomes independent in another.

In the goods-producing (manufacturing) sector, the differential

shift in town j between time periods (o and t) is expressed as a function of:

(1) the sum of the D.S. in all non-goods producing sectors in that time interval in that town

(2) the total goods producing sector employment in that town at time = 0

(3) the total non-goods producing sector employment in that town at time = 0

(4) the holding capacity for manufacturing employment in that town at time = 0

(5) the index of accessibility (A_{ij}) of that town to employment in Business and Professional Services at time = 0.

where: $$A_{ij} = \sum_{j=1}^{n} E_{ik} d_{iv}^{a} \qquad (4)$$

with: E_{ik} = level of activity i in town k

d_{iv} = shortest travel time on highway network between towns j and v

a = an empirically determined exponent.

Presentation of the manufacturing sector model equation will be sufficient to demonstrate the general model structure and format.

$$D_j^m = a_1 \sum_k D_{kj}^s + b_1 E_{jo}^m + b_2 \sum_k E_{kjo}^s + b_3 A_{jo}^B + b_4 H_o^m + b_o \qquad (5)$$

where: E = level of activity

D = differential shift

A = potential or accessibility of a town to an activity

H = holding capacity for activities (maximum level restricted by policy)

M = manufacturing sector

S = services sector

B = business and professional service sector

P = population

k = industry group in the employment sector

m = number of industry groups (k = 1,2,...,6)

j = town in the state

o = beginning of growth period

t = end of growth period

a,b = parameters

These estimated values of differential shift then form the basic measure of economic growth of towns in Connecticut used in the present study.

VII. 6. H. d. Data Collection

Data collection focused on the delineation of the historical pattern of population and employment growth by towns in the state. Employment data by town by two-digit SIC category were collected for 1950 and 1962, mainly from the records of the State Department of Employment Security. The income distribution by towns was obtained from census tapes and records. The Connecticut Development Commission developed the holding capacity estimates for population and manufacturing employment. The various accessibility measures were computed on a highway network, specifically prepared for 1950 by the Highway Department.

Review of the above mentioned data requirements substantiates, at least partially, the operational simplicity of the present model under study. Because much of the necessary data should presently exist in readily available form, relatively small expenditures of time and money would be required for their acquisition. One can only truly appreciate this factor after he has reviewed some of the more costly and complex allocation models (discussed later in this report).

VII. 6. H. e. Employment and Population Projections

Using some form of parameter estimation procedure (e.g. single equation, two-stage least squares, etc.) the structural coefficients are determined. In general, the relative signs and magnitude of the Connecticut coefficients conform to *a priori* expectations. For most equations a high order of reliability was expressed by the R^2 and S.E.E. values obtained.

The projection procedure for each iteration period (10-year increments from 1960 - 2000) involves the following steps:

(1) Projection of Differential Shifts (D_{ij}) The values of the independent variables in each equation for each town for the beginning of the growth period are multiplied by their coefficients in the reduced-form equations (solved for one dependent variable).

(2) Computation of Proportional Shift (P_{ij}) The projected state activity growth rates are applied to the level of the corresponding activity in each town at the beginning of the growth period to get the P_{ij} for each town.

7200-8-SB
5-50
B-C---T

(3) Computation of Net Change $(N_{ij}) = D_{ij} + P_{ij}$.

(4) Holding Capacity Checks Net changes are then compared with maximum additional holding capacities for each activity in each town.

(5) Unique Locations "Hand allocation" of the employment of these "unique" activities which cannot be accurately described by the equation system.

(6) Recursive Projections Update the values of the independent variables to correspond with the beginning of the next growth period.

VII. 6. H. f. Conclusion

In consideration of the preceding discussion, the Voorhees model is both theoretically and operationally appealing. Comparisons of the predicted differential shifts with those actually observed between 1950 and 1962 were generally quite acceptable. The simplicity of the data requirements, model formulation, and manipulation warrant its consideration for incorporation into any state projection and allocation study.

VII. 7. Additional Models

Numerous other allocation models were examined but, because of their extreme complexity and large investment requirement, were not felt to be appropriate, and therefore have not been summarized in this report.

While it is difficult to identify the specific costs for and particular model, a rough estimate for a full-scale land use model is probably at least $500,000. Such an investment is due to the

comprehensive nature of such models. The overly comprehensive structure of existing large-scale models has two aspects: (1) the models were designed to replicate too complex a system in a single shot; and, (2) they were expected to serve too many purposes at the same time. Hypercomprehensiveness usually means too many variables and too much detail are included in the model structure. This is due in part to the desire of model-builders to attempt to capture the complex interractions of the many diverse factors which shape urban land use patterns. Obviously such attempts necessitate the inclusion of large number of variables into the models, the collection or generation of data which these variables represent, and long hours of statistical analysis to determine whether such variables are significant, and if so, to what degree.

Such modeling attempts can feasibly only be undertaken by the largest city or regional planning agencies, and even then they must be assisted by significant amounts of federal funding. The technical expertise required to effectively operate and interpret the results of large scale models is for the most part unavailable at the smaller agency level.

Even if the existing large-scale models did produce results superior to the simpler techniques, and even if they were in an operational state where they could be applied to areas other than those which they were developed, problems still exist. Application of such a model would obviously require that it be calibrated for

the area under study. Depending upon the particular model, this would necessitate the generation of such data as development costs per parcel, rents per unit of land, and complex location probabilities, all at a considerable expenditure of money and effort. Thus even those models purported to be applicable to any area can only be applied after a substantial amount of new data has been generated.

Finally, these large-scale models have been excluded from discussion because they are essentially modifications of the simpler techniques reviewed in this report. For example, the models of the Bay Area Simulation and Transportation Studies - BASS I, BASS III, and PLUM - are all modifications to a degree of the basic "Lowry Model." In addition, the BASS models incorporate a regression analysis methodology of the "weighted overlay procedure". Thus the kinds of outputs generated by such models is very similar to that of the simpler techniques. The parsimonious nature of these simpler models can be better appreciated in light of the extreme data and expenditure requirements of the complex models. The PLUM model, for example, requires 38 pages of programming statements in order to analyze the vast amount of variables and transformations incorporated into the model structure.

Although not discussed separately, the gravity or interaction model concept is inherent in the structure of most of the reviewed models. The simplest form of gravity model is derived directly from the Newtonian analogy that the force ($F_{1,2}$) which two masses (M_1,M_2)

exert upon each other is proportional to the product of the masses and inversely related to the distance between them ($d_{1,2}$):

$$F_{1,2} = K \frac{M_1, M_2}{d^a_{1,2}} \qquad (1)$$

where K and a are predetermined constants.

The concept can be utilized to describe the intensity of interaction ($G_{1,2}$) between area 1 and area 2 as being a function of the relative attractiveness, size etc. (W_1, W_2) of each area and the distance between them.

$$G_{1,2} = K \frac{W_1, W_2}{d^a_{1,2}} \qquad (2)$$

This is the basic procedure used by most models to allocate residential growth (P_j) to area j, based upon the employment in each area within the study region and the time, cost of travel separating j and other i areas ($T_{i,j}$).

$$P_j = \sum_{i=1}^{N} \frac{E_i}{T^a_{ij}} \qquad (3)$$

In the above formulation, employment in each area replaces the mass or attractiveness variables in the preceeding equations. The procedure is further illustrated in several of the earlier reviewed techniques. There are several further modifications of the basic gravity concept which are possible, such as the imposition of alternative types of constraints, and the interested reader is referred to any of several excellent texts, some of which are listed in the bibliography to this report, on the concept of interaction models (Lanchester Polytechnic

1969) (U.S. Dept. of Commerce 1963).

VII, 8. Conclusion

When one undertakes an investigation of land use allocation models one soon becomes overwhelmed by the number of models in existence and the vast amount of literature either eulogizing or disparaging the results. Much of the criticism receives its support from the complexity of the processes of urban land use location which the models attempt to approximate. Proximity to employment, personal income, land prices, business profit-maximization, and numerous other factors all effect the spatial organization of urban activities. Thus it is little wonder that any attempt to simulate the interactions of such factors will draw a considerable expression from both proponents and opponents.

The models discussed in this report represent a sample of the many simpler, more operational models. It is felt by the author that such models possess the potential of producing reasonable, accurate and acceptable estimations, subject to their particular limitations and biases. Whether one uses the complex, large-scale models or the smaller, simpler techniques, one thought must be kept closely in mind. The results of any model are somewhat tenuous. They represent only approximations of what the future land use pattern may be, and as such should be interpreted and utilized with discretion. No model can perfectly mirror actual conditions. Thus any plans based upon the results of such models should allow for the inclusion of staff judgemental modification. Any model estimations of future

locational patterns must be interpreted in light of the knowledge and experience of local officials, planners, civic-leaders, and others possesses valuable potential input into the locational analysis.

VIII. Summary

The preceeding report has presented a literature review and methodology evaluation of several techniques of land use planning analysis. The concepts of projecting population and economic activity growth, and of spatially allocating components of that growth to areal sites are of considerable importance whether one is concerned with providing the required services and facilities to sustain such growth, or whether the center of interest is air quality maintenance planning.

The importance of making and allocating projections for AQMP is discussed in detail in Volume 4 of the Guidelines for Air Quality Maintenance Planning and Analysis, and need not be reiterated here. Identification of future emission levels and the development of effective strategies to control such emissions requires that estimates be made of: (1) the expected increases in the sources of these emissions; and, (2) the spatial location of these sources.

The techniques contained in this report represent a wide range of alternatives for making such estimations, from the simpler techniques such as economic base analysis and the weighted overlay procedure, to the more complex such as the economic-ecologic linkages models. Regardless of the procedure(s) selected for use, the derived model outputs must be interpreted in light of the limitations and assumptions upon which the estimates are based. The present state-of-the-art of projecting and allocating urban growth is far from perfect. However, familiarity on the analyst's part with the factors and problems affecting

urban growth and location can do much to mitigate many of the present inadequacies of such techniques. Model estimates must always be tempered with sound judgement and consideration of the unquantifiable determinants of the population and economic growth.

REFERENCES

Bartholomew, H. and Assoc. (1970). South Bend Urban Area Transportation Study. South Bend, Indiana.

Bonner, E.R. and Fahle, V.L. (1967). Technique for Area Planning. Regional Economic Development Institute, Pittsburgh.

Center for Real Estate and Urban Economics (1968). Jobs, People, and Land: Bay Area Simulation Study (BASS). U. California, Berkeley.

Chapin, F.S. Jr. (1965). Urban Land Use Planning. U. Illinois, Urbana.

Chapin, F.S. Jr. (1965). "A Model for Simulating Residential Development," Journal American Institute Planners.

Chicago Area Transportation Study (1960). Final Report.

Connecticut Highway Department (1961). Hartford Area Traffic Study.

Doekson, G.A. and Little, C.H. (1968). "Effects of the Size of the Input-Output Model on the Results of an Impact Analysis," Agricultural Economics Research.

Glickman, N.J. (1971). "An Econometric Forecasting Model for the Philadelphia Region," Journal of Regional Science.

Goldberg, M. (1969). Intrametropolitan Industrial Location, U. California, Berkeley.

Goldner, W. (1971). "The Lowry Model Heritage," Journal American Institute Planners.

Hall, O.P. and Licari, J.A. (1974). "Building Small Region Econometric Models: Extension of Glickman's Structure to Los Angeles," Journal Regional Science.

Hansen, W.L. and Tiebout, C.M. (1963). "An Intersectoral Flows Analysis of the California Economy," Review Economics and Statistics.

Hewings, G.J.D. (1972). "Aggregation for Regional Impact Analysis," Growth and Change.

Hill, D.M. (1965). "A Growth Allocation Model for the Boston Region," Journal American Institute Planners.

Hite, J.C. and Laurent, E.A. (1972). Environmental Planning: An Economic Analysis. Praeger.

Ingram, G.K. et.al. (1972). The Detroit Rototype of the NBER Urban Simulation Model. Columbia U.P.

Isard, W. (1960). Methods of Regional Analysis. MIT Press.

Isard, W. et. al. (1972). Ecologic - Economic Analysis for Regional Development. Free Press.

James, F.J. (ed) (1974). Models of Employment Location. Center Urban Planning Research, Rutgers University, New Brunswick, N.J.

Jhaveri, D. and Al Chalabi, S. (1973). Activity Allocation Land Use Model for the Northeastern Illinois Region. Northeastern Illinois Planning Commission, Research Memorandum 18a.

Klein, L.R. (1969). "The Specification of Regional Econometric Models," Papers Regional Science Association.

Kohn, R.E. (1974). "Industrial Location and Air Pollution Abatement," Journal Regional Science.

Korbach, R.J. (1972). An Environmental Linkages Model of the State of Maryland. Bureau Business and Economic Research, University Maryland, College Park.

Ladshmanan, T.R. and Hansen, W.G. (1965). "A Retail Market Potential Model," Journal American Institute Planners.

Lanchester Polytechnic (1969). Gravity Models in Town Planning. Coventry, England.

Lathrop, G.T. and Hamburg, J.R. (1965). "An Opportunity - Accessibility Model for Allocating Regional Growth," Journal American Institute Planners.

Leigh, R. (1970). "The Use of Location Quotients in Urban Economic Base Studies," Land Economics.

Lowry, I.S. (1964). A Model of Metropolis. Rand Corp.

Lowry, I.S. (1964). Seven Models of Urban Development. Rand Corp.

McHarg, I. (1972). Design with Nature. Natural History Press.

Masser, I. (1972). Analytical Models for Urban and Regional Planning. Wiley.

Massey, D. (1969). Some Simple Models for Distributing Changes in Employment within Regions. Center for Environmental Studies, Working Paper No. 24.

Mathur, V. and Rosen, H. (1974). "Regional Employment Multiplier: A New Approach," Land Economics.

Miernyk, W.H. et.al. (1970). Simulating Regional Economic Development. Heath - Lexington.

Miernyk, W.H. and Seers, J.T. (1974). Air Pollution Abatement and Regional Economic Development. Heath - Lexington.

Perloff, H.S. et.al. (1961). Resources and Economic Growth. Johns Hopkins.

Pfouts, R.W. (ed) (1970). The Techniques of Urban Economic Analysis. Chandler - Davis.

Putnam, S.H. (1967). "Intraurban Industrial Location Model: Design and Implementation," Papers Regional Science Association.

Putnam, S.H. (1970). "Developing and Testing an Intra-Regional Model," Regional Studies.

Reif, B. (1973). Models in Urban and Regional Planning. Intertext Publishers.

Schlager, K.J. (1965). "A Land Use Plan Design Model," Journal American Institute Planners.

Tiebout, C.M. (1962). The Community Economic Base Study CED. Supplementary Paper No. 16.

Traffic Research Corporation (1967). EMPIRIC Land Use Forecasting Model.

U.S. Department of Commerce (1963). Calibrating and Testing a Gravity Model for any Size Urban Area.

Vorhees, A.M. and Assoc. (1966). A Model for Allocating Economic Activities into Sub-Areas in a State.

Williamson, R.B. (1970). "Simple Input-Output Models for Area Economic Analysis," Land Economics.

Wilson, A.G. (ed). Urban and Regional Planning. Pion, London.

UNIVERSITY OF ILLINOIS-URBANA
3 0112 062889495